8-9-2019

Nuclear Organization and the Epigenetic Landscape of the Mus musculus X-Chromosome

Alicia Liu
University of Connecticut - Storrs, alicia.liu@uconn.edu

Nuclear Organization and the Epigenetic Landscape of the *Mus musculus* X-Chromosome

Alicia J. Liu, Ph.D.
University of Connecticut, 2019

ABSTRACT

X-linked imprinted genes have been hypothesized to contribute parent-of-origin influences on social cognition. A cluster of imprinted genes *Xlr3b*, *Xlr4b*, and *Xlr4c*, implicated in cognitive defects, are maternally expressed and paternally silent in the murine brain. These genes defy classic mechanisms of autosomal imprinting, suggesting a novel method of imprinted gene regulation. Using *Xlr3b* and *Xlr4c* as bait, this study uses 4C-Seq on neonatal whole brain of a 39,XO mouse model, to provide the first in-depth analysis of chromatin dynamics surrounding an imprinted locus on the X-chromosome. Significant differences in long-range contacts exist between X^M and X^P monosomic samples. In addition, X^M interaction profiles contact a greater number of genes linked to cognitive impairment, abnormality of the nervous system, and abnormality of higher mental function.

This is not a pattern that is unique to the imprinted *Xlr3/4* locus. Additional

4C-Seq experiments show that other genes on the X-chromosome, implicated in intellectual disability and/or ASD, also produce more maternal contacts to other X-linked genes linked to cognitive impairment. The observation that there are differences between the maternal and paternal X interactomes is bolstered by potential variation in Atrx binding and H3K27me3 enrichment between X^M and X^P, suggesting that there may be broad-scale differences of the X-chromosome, related to parent-of-origin effects. Taken together, these results provide intriguing insight into the maternal X susceptibility to cognitive and social impairment.

Nuclear Organization and the Epigenetic Landscape of the *Mus musculus* X-Chromosome

Alicia J. Liu

Bachelor of Science, University of Delaware, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Nuclear Organization and the Epigenetic Landscape of the *Mus musculus* X-Chromosome

Presented by
Alicia J. Liu, B.S.

Major Advisor _____
Michael O'Neill

Associate Advisor _____
Rachel O'Neill

Associate Advisor _____
John Malone

University of Connecticut
2019

To Fat Stanley, protector of the pumpkin patch.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to the people in my life whose encouragement and support made this dissertation possible.

First and foremost, the biggest thanks go to Mike. Thank you for being my advisor. Thank you for inspiring me and motivating me to meet every challenge, while also giving me the independence to grow and mature into the scientist I am today. The skills I learned in your lab I will carry with me the rest of my life.

Besides my advisor, I extend thanks to the rest of my committee members for their years of support and invaluable advice. Rachel and John, your insights have been an incredible resource to me over the course of this dissertation. Leighton and Bibi, your course work and expertise provided the foundation upon which a lot of this work was built. What I have learned from my committee has been a tremendous help in designing and analyzing experiments, and for that I am incredibly grateful.

I would also like to thank my lab mates, both past and present, for their continued support, and for making the M. O'Neill lab a wonderful, fun, and productive environment. Natali, for always being there as a great friend and supportive brain to lean on. Rob, who answered my endless questions and was always willing to lend a helping hand. And to Glenn, Matt, Katelyn, and Tyler, thank you.

Special thanks go to Dylan. You are my motivation and my inspiration. Your unwavering faith in me made this possible.

Last but not least, I would like to thank my friends and family for their constant encouragement, love, and support throughout my long graduate career.

Thank you.

TABLE OF CONTENTS

Introduction : Setting the Problem 1
0.1. Overview 1
0.2. Genomic Imprinting 3
0.3. Eutherian X-Chromosome Dosage Compensation 5
0.4. Turner Syndrome 6
0.5. Autism Spectrum Disorder 8
0.6. X-linked Genomic Imprinting 10
0.7. Mechanisms of Epigenetic and Imprinting Regulation 12
0.7.1. Differential DNA Methylation 13
0.7.2. Allele-Specific Histone Modification 14
0.7.3. Interactome Analyses of Imprinted Domains 16

Ch. 1 : Nuclear Organization of the Imprinted *Xlr3/4* Locus 18
1.1. Abstract 18
1.2. Background 19
1.3. Results 24
1.3.1. Generating a library of genome-wide chromatin interactions 24
1.3.2. Long-range interactions at the imprinted *Xlr3/4* locus 25
1.3.3. 4C domains reside in open chromosomal compartments 30
1.3.4. Short-range comprehensive analysis of looping interactions 33
1.3.5. 4C domains enriched with genes linked to cognitive impairment 40
1.4. Discussion 45

Ch. 2 : Epigenomic Profiling of the Maternal and Paternal X-Chromosomes . . 48
2.1. Abstract 48
2.2. Background 49
2.3. Results 53
2.3.1. Interaction profiles of the maternal and paternal X-chromosomes 53
2.3.2. X^M 4C domains are enriched with genes involved in cognitive function . 55
2.3.3. X-chromosome profiling of H3K27me3 and Atrx 58
2.4. Discussion 68

Ch. 3 : Regulation of *Xlr3b* Gene Expression 73
3.1. Background 73
3.2. Results 78
3.3. Discussion 83

Ch. 4 : Synthesis and Future Directions 86

Ch. 5 : Materials and Methods 93
5.1. Animal Breeding and Tissue Collection 93
5.2. *Xlr3b* Expression Analysis 93
5.3. Chromosome Conformation Capture (3C) 94
5.3.1. Mouse Neonatal Brain Tissue 94
5.3.2. Mouse Fibroblast Cultured Cells 96
5.4. 3C-Carbon Copy (5C-Seq) 97
5.4.1. 3C Control Library Generation 97
5.4.2. 5C Primers and Library Generation 97
5.4.3. Illumina MiSeq Sequencing 98
5.4.4. Data Analysis 98
5.5. Circular 3C (4C-Seq) 99
5.5.1. 4C Primers and Library Generation 99
5.5.2. Illumina NextSeq500 Sequencing 100
5.5.3. Data Analysis 101
5.6. Chromatin Immunoprecipitation (ChIP-Seq) 102
5.6.1. ChIP Library Generation 102
5.6.2. Illumina NextSeq500 Sequencing 103
5.6.3. Data Analysis 103
5.7. Assay for Transposase-Accessible Chromatin (ATAC-Seq) 103

App. A : 104
A.1. Primers 104
A.2. BACs 118
A.3. Antibodies 118

App. B : 119
B.1. Sequencing Metrics 119

Bibliography 125

LIST OF FIGURES

0.4.1 Cognitive defects in individuals with Turner syndrome. 7
0.6.2 Genomic imprinting of *Xlr3b, Xlr4b*, and *Xlr4c*. 11
0.6.3 Schematic representation of the *Xlr3/4/5* locus. 12
0.7.4 Increased *Xlr3b* expression in 5-aza-2-dC treated cell lines. 15

1.2.1 Overview of 3C-derived methods. 21
1.3.1 Overview of 4C-Seq bait region at the imprinted *Xlr3/4* locus. 25
1.3.2 Distribution of 4C-Seq reads. 26
1.3.3 Distribution of 4C interactions. 27
1.3.4 Significant 4C domains in *cis* and in *trans*. 28
1.3.5 Long-range *cis* interactions of *Xlr3b* and *Xlr4c*. 29
1.3.6 DNA replication timing of 4C domains. 31
1.3.7 Enrichment of active chromatin signatures around 4C domains. 32
1.3.8 Repetitive elements within 4C domains. 34
1.3.9 Near-bait interactions of *Xlr3b* and *Xlr4c*. 35
1.3.10 Near-bait interactions show parent-specific differences within a TAD. 36
1.3.11 Overview of 5C-Seq primer design at the imprinted *Xlr3/4* locus. 37
1.3.12 5C interaction profile of *Xlr3b*. 38
1.3.13 4C interacting regions display characteristics of regulatory elements. 39
1.3.14 Distribution of region-gene associations within 4C domains. 41
1.3.15 Functional annotation of genes within *Xlr4c* 4C domains. 42
1.3.16 Functional annotation of genes within *Xlr3b* 4C domains. 44

2.3.1 Significant 4C domains of six viewpoints across the X-chromosome. 54
2.3.2 Number of associated genes per 4C domain. 57
2.3.3 Distribution of region-gene associations within 4C domains. 58
2.3.4 Functional annotation of genes within 4C domains. 59
2.3.5 Human phenotype functional annotation of 4C domains. 60
2.3.5 Human phenotype functional annotation of 4C domains *(cont.)* 61
2.3.6 Motifs within X^M 4C domains. 62
2.3.7 Heatmap of Atrx and H3K27me3 ChIP-Seq peaks to TSS regions. 63
2.3.8 Heatmap of ATAC-Seq peaks to TSS regions. 64
2.3.9 Average profile of ATAC-Seq and ChIP-Seq peaks to TSS regions. 65
2.3.10 Genomic annotation of Atrx binding sites on the X-chromosome. 66
2.3.11 Genomic annotation of H3K27me3 peaks on the X-chromosome. 67
2.3.12 Atrx KEGG enrichment analysis. 68
2.3.13 H3K27me3 KEGG enrichment analysis. 69

3.2.1 RNA pol II enrichment at the 5′ end of *Xlr3b* 79

3.2.2 RNA pol II enrichment across *Xlr3b* 80
3.2.3 Predicted G-quadruplex structures at *Xlr3b*. 81
3.2.4 Atrx enrichment at the *Xlr3b* TSS. 82
3.2.5 4C Primer Design in *Xlr3b*. 82
3.2.6 CTCF signal across *Xlr3b*. 83
3.2.7 CTCF enrichment at *Xlr3a* and *Xlr3b*. 84

5.1.1 Breeding scheme for the generation of X-monosomic mice. 94

LIST OF TABLES

1.3.1 Unique *cis* interactions of *Xlr4c* 43

2.3.2 Enrichment of cognitive function annotations 56

A.1.3 General Primers 104
A.1.4 5C Primers 106
A.1.5 4C Primers 116
A.2.6 BAC Clones 118
A.3.7 ChIP Antibodies 118

B.1.8 5C-Seq Library Metrics 119
B.1.9 4C-Seq Library Metrics 119
B.1.10 ChIP-Seq Library Metrics 124

Introduction

Setting the Problem

0.1 Overview

The term epigenetics describes heritable modifications to chromatin or DNA, that are known to regulate the transcriptional activity of the underlying DNA sequence. These modifications allow the genome to regulate gene expression in a time and tissue-specific manner, and can be influenced by age, environment, lifestyle, as well as disease state [1]. The most well-characterized epigenetic marks are DNA methylation and histone protein modifications. Methylated promoters typically result in gene silencing [2]. Post-translational modifications to histone proteins can alter chromatin structure, resulting in transcriptional activation or repression [3]. Chromatin remodeling is a dynamic change in chromatin architecture, going from a closed, condensed state to an open, accessible state, thereby controlling gene expression. Two mammalian-specific phenomena regulated by epigenetic factors are X-chromosome inactivation (XCI) and genomic imprinting.

XCI is a mechanism of dosage compensation to transcriptionally silence one of the two X-chromosomes in females. This silencing emanates from the X-inactivation center (XIC) and involves the coating of a non-coding RNA across the chromosome

to be inactivated, creating a repressed, heterochromatic environment [4]. XCI can either be imprinted, in which the paternal X-chromosome is selectively inactivated, or random.

Genomic imprinting describes parent-of-origin dependent gene expression, and it relies on the DNA being biochemically marked with information about its parental origin [5]. These epigenetic marks typically include differentially methylated DNA and allele-specific histone modifications (ASHM). Recent studies have shown that H3K27me3 serves as an imprinting mark for DNA methylation-independent autosomal imprinting and *Xist* for paternally imprinted XCI [6][7]. This led us to ask whether it is also responsible for the imprinting of a cluster of paternally silenced genes on the mouse X-chromosome, and whether there are differences chromosome-wide related to X-linked genomic imprinting.

Since their imprinted state was established in 2005, no differentially methylated region (DMR) has been found surrounding the imprinted genes: *Xlr3b, Xlr4b*, and *Xlr4c*. This locus is the only imprinted cluster found on the mouse X-chromosome, and the mechanism behind their imprinted expression remains elusive. The following work aims to characterize the nuclear organization of the mammalian X-chromosome and determine if differences exist between maternal and paternal X-chromosomes in regards to chromatin interactions and accessibility. This body of work presents the following results as they pertain to parent-of-origin-related differences on the mouse X-chromosome:

1. Local interactions within a 2.7 Mb region surrounding the imprinted *Xlr3/4*

locus show the most dramatic differences and likely reflect a mechanism that controls imprinted gene expression.

2. In neonatal mouse brain, the maternal X-chromosome interacts with more genes related to intellectual disability and cognitive impairment, which is possibly related to the susceptibility to cognitive and social impairment found in individuals who solely inherit a maternal X-chromosome.

3. Genomic annotation of H3K27me3 and Atrx ChIP-Seq peaks in neonatal mouse brain, reveals differences between maternal and paternal X-chromosomes.

Evidence shown here suggests that there are fundamental differences between the maternal and paternal X-chromosomes occurring above the level of DNA. Such differences might play a role in the parent-of-origin effects in Turner syndrome and in the male prevalence of Autism Spectrum Disorder (ASD) [8][9]. Ultimately, I hypothesize that the interactomic differences between the maternal and paternal X-chromosomes could be related to a maternal X susceptibility to intellectual disability and cognitive impairment.

0.2 Genomic Imprinting

Genomic imprinting is differential gene expression based on the parent-of-origin of the alleles inherited. Imprinted genes are epigenetically marked in gametes, through a process that involves erasing existing marks and rewriting them based on the parental genome [10]. For diploid genes, though we inherit two copies, genomic imprinting leads to exclusive expression of genes from only one parent. Though

imprinting negates the benefits of diploidly, proper imprinting is fundamental for normal development. Many imprinted genes are shown to be active during key developmental time points, though there are some that maintain their imprinted status into adulthood [5]. As many imprinted genes are associated with cell proliferation and fetal growth, imprinting dysregulation has been implicated in cancer and tumor development [11]. Classic imprinting disorders include Prader-Willi and Angelman syndromes, Silver-Russell syndrome, and Beckwith-Wiedemann syndrome. The incidence of these and other imprinting disorders are increased with the use of assisted reproductive technology (ART) [12]. Thus, individuals conceived through *in vitro* manipulation of gametes may be at a higher risk of imprinting dysregulation.

It has long been known that the maternal and paternal genomes are not equivalent. Early experiments by Takagi and Sasaki demonstrated preferential inactivation of the paternal X-chromosome in extra-embryonic tissues of the developing mouse [13]. About a decade later, Surani and Barton established that mouse parthenogenetic (PG) embryos fail to develop to term, suggesting that specific imprinting must occur during gametogenesis, such that both a maternal and paternal pronucleus are required for proper development [14].

Haig's Kinship theory for genomic imprinting proposes that imprinting evolved in mammals because of an evolutionary conflict between maternal and paternal genomes, representing a form of intragenomic conflict [15]. This theory argues that in viviparous organisms, paternal alleles demand more maternal resources, in the interest of their own reproductive success, as females can produce offspring from

more than one male. However, the cost of growth of the offspring falls predominantly on the mother. This conflict over the distribution of maternal resources suggests that paternally expressed transcripts would act as enhancers of prenatal growth and maternally expressed transcripts would act as inhibitors of prenatal growth. While some imprinted loci fall into the lines of the Kinship theory, there are many imprinted genes whose functions don't fit the conflict model or still remain a mystery.

0.3 Eutherian X-Chromosome Dosage Compensation

Genomic imprinting shares many mechanistic similarities with X-chromosome inactivation, a phenomenon that takes place in eutherians, or placental mammals, and marsupials [16]. In eutherian females, one of the two X-chromosomes is randomly inactivated early in development, and this inactive state is maintained in all subsequent somatic cell divisions. However, in extra-embryonic tissues, ie. the placenta, and in marsupials, we see imprinted X inactivation, where the paternal X-chromosome is always inactivated over the maternal counterpart. Though random XCI was discovered first, it is believed that imprinted XCI evolved first [17].

It has been theorized that genomic imprinting and XCI co-evolved after separation from monotremes, or egg-laying mammals. With the emergence of the placenta, came selective pressure for the imprinting of growth-regulating genes, as stated by the Kinship theory. Imprinted XCI evolved as the primary mechanism for dosage compensation, with histone modifications being the primary epigenetic marks. Random XCI came later in the somatic tissues of eutherian mammals as an evolutionary

innovation that confers mosaicism, along with the benefits of heterozygosity. DNA methylation was further adopted for the stability of the imprint [18].

Mechanistic parallels serve as additional evidence for the co-evolution of genomic imprinting and XCI. Beyond DNA methylation and histone modifications, both phenomena utilize the expression of antisense non-coding transcripts, and are correlated with asynchronous DNA replication and chromatin condensation [19]. At a mechanistic level, we have two processes that produce monoallelic expression through similar means, which is suggestive of an evolutionary link that forms the foundation of this research.

0.4 Turner Syndrome

Monosomy of the X-chromosome in human females is a genetic condition described as Turner syndrome, the severity of which is highly dependent on the extent of monosomy. Characteristics of Turner syndrome include short stature, gonadal dysgenesis, dysmorphic stigmata, and abnormalities in kidney and heart. Patients are also reported to have a high stress tolerance, a tendency towards overcompliance, limitations in emotional competence, and selective impairments in non-verbal, visual-spatial information processing [20]. The wide spectrum of phenotypes seen in Turner syndrome can be attributed to haploinsufficiency of genes expressed in both sex chromosomes, such as genes located in the pseudoautosomal region (PAR), and possibly, to genomic imprinting of X-linked genes.

In 1997, Skuse and colleagues published data analyzing social cognition in pa-

tients with Turner syndrome [8]. They observed differences in the severity of symptoms based on the parent-of-origin of the single X-chromosome, and hypothesized that these differences may be due to an X-linked imprinted locus. In this study, they found that 45,X^P females had superior verbal intelligence and higher-order executive function skills, in comparison to 45,X^M females (Figure 0.4.1). Normal males performed similarly worse in comparison to normal females. This data, along with the observed male bias in ASD, altogether suggests that there may exist genes on the X-chromosome, subject to genomic imprinting, that impair social cognition in individuals that solely inherit a maternal X-chromosome.

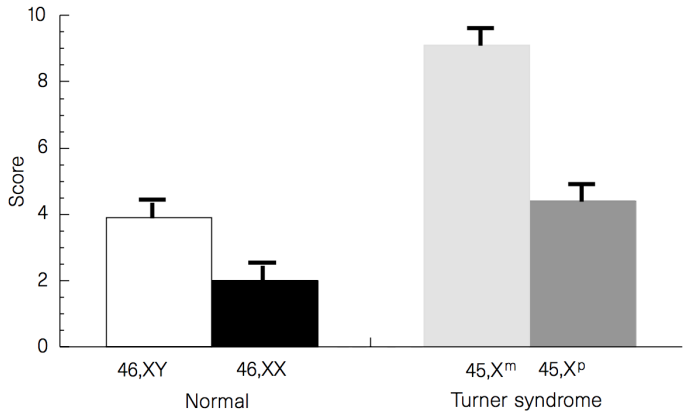


Figure 0.4.1: **Cognitive defects in individuals with Turner syndrome.** Bar graph of average scores on test of social cognitive impairment. The higher the score the lower the social cognitive ability demonstrated on the test. Originally published by Skuse *et al.* [8].

A study by Lepage, investigating regional cortical volume, thickness, and surface area in females with Turner syndrome hoped to explain the variability in cognitive and behavioral phenotypes with regards to the parental origin of the single X-chromosome [21]. They found that 45,X^P females demonstrated thicker cortex in

the temporal regions bilaterally in comparison to 45,X^M, and 45,X^M females showed bilateral enlargement of gray matter volume in the superior frontal regions in comparison to 45,X^P. These differences in cortical thickness and volume, dependent on the parental origin of the X-chromosome in Turner patients, further suggest that X-linked imprinted genes play a role in early brain development, manifesting into the social-cognitive phenotypes seen in these patients.

The X-chromosome is highly enriched for genes related to intelligence and neurocognition [22]. The neurocognitive phenotype of Turner syndrome is a complex trait, likely influenced by haploinsufficiency of multiple genes, each contributing to the phenotypic variance seen in these patients [23]. Turner syndrome candidate genes include those that escape X inactivation, and imprinted genes on the X-chromosome.

0.5 Autism Spectrum Disorder

X-linked genomic imprinting has also been implicated in the male prevalence of autism spectrum disorder (ASD), affecting males to females at a ~4.3:1 ratio [9]. ASD is a complex developmental disorder that affects social communication and behavior. Epidemiological and genetic studies suggest that the sex bias of ASD may be due, in part, to a female protective effect (FPE) that results in a higher diagnosis threshold and a reduced incidence of ASD in females [24]. X-chromosome mosaicism resulting from random X inactivation may provide some of this defense against recessive X-linked mutations. However, studies have not found a single

genetic locus that mediates the FPE, though it is possible that multiple genetic loci play a role in the sex bias of ASD.

Studies of Turner syndrome patients show not only do 45,X^M females perform worse on social cognition tests than 45,X^P females, but they are also more vulnerable to ASD [25]. Imprinted genes tend to be highly expressed in the brain, showing dynamic variability in when and where they are expressed. This is consistent with neurodevelopmental defects seen in many imprinting disorders [26]. Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are two autosomal imprinting disorders known to cause neurological impairment and behavioral phenotypes. PWS results when the paternal allele of the 15q11-q13 locus is lost, and is associated with behavioral and social impairment, along with higher rates of ASD [27]. Whereas AS results when the maternal allele is lost, and is associated with severe cognitive and neurological impairment [28].

Mouse chimera experiments offer additional evidence for the essential role that imprinted genes play in neurodevelopment. These experiments have shown not only that parthenogenetic (PG) embryos, with only a maternally inherited genome, and androgenetic (AG) embryos, with only a paternally inherited genome, fail to develop to term, but also that PG and AG cells localize to different regions in the mouse brain during early development [29]. PG cells accumulate in the cortex, striatum, and hippocampus, while AG cells are enriched in the hypothalamus and pre-optic areas, suggesting that imprinted genes can have specific effects on brain size, organization, and development [30].

Linkage studies have shown that the severity of ASD, as well as schizophrenia and bipolar disorder, may depend on the sex of the transmitting parent, further implicating parent-of-origin effects [31]. The prevalence of ASD in males, who are hemizygous for the X-chromosome, supports an X-linked imprinted locus in a multifactorial threshold model involving the cumulative effect of many genes.

0.6 X-linked Genomic Imprinting

Using a mouse model of Turner syndrome, Raefski and O'Neill identified a cluster of paternally imprinted genes on the mouse X-chromosome [32]. *Xlr3b, Xlr4b*, and *Xlr4c* are maternally expressed and paternally silent in mouse brain and fibroblast (Figure 0.6.2). Simultaneously, Davies *et al.* published data showing parent-of-origin effects on cognitive function using the same Turner syndrome mouse model [33]. They demonstrated that 39,X^M mice show deficits in Y-maze reversal learning relative to 40,XX and 39,X^P subjects.

Two more X-linked imprinted genes were identified by Kobayashi *et al.* in the following few years [34][35]. Using DNA microarrays to compare gene expression between male and female mouse blastocysts, they showed that *Rhox5/Pem* is predominately expressed in female blastocysts, showing preferential expression from the paternal X-chromosome. However, at embryonic day 7.5, the imprint switches, and *Rhox5/Pem* is predominantly expressed from the maternal allele. Later, they showed that *Fthl17* displays preferential paternal X-chromosome expression in female blastocysts as early as the two-cell stage, and is no longer detectable by embryonic day

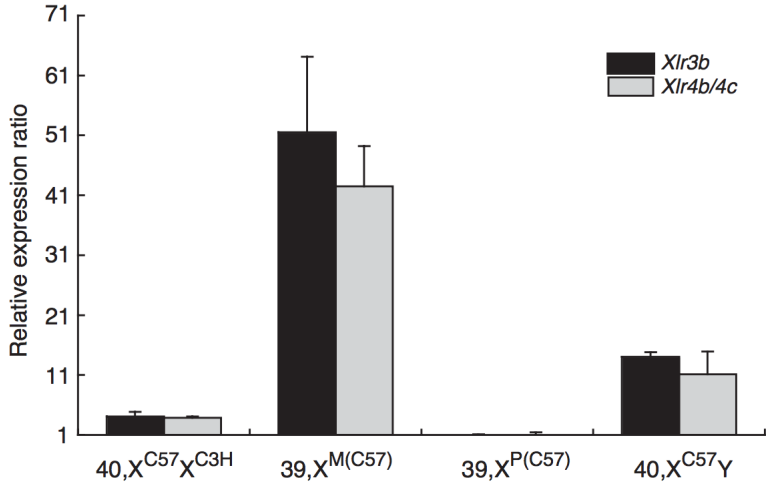


Figure 0.6.2: **Genomic imprinting of *Xlr3b, Xlr4b,* and *Xlr4c*.** Real-Time PCR showing imprinted expression of three genes in 39,X^M and 39,X^P neonatal brain, normalized to *ActB*. $n = 3$ for each sample type. Originally published by Raefski and O'Neill [32].

9.5. Unlike *Rhox5/Pem* and *Fthl17*, the *Xlr3b/4b/4c* locus is paternally imprinted, and holds this imprinted expression from as early as the eight-cell stage up to adult.

Like *Rhox5/Pem* and *Fthl17*, the *X-linked lymphocyte-regulated (Xlr)* genes belong to an expanded, multi-copy gene family on the X-chromosome. The *Xlr* locus exists as two clusters, each containing multiple duplications of a triad in the orientation *Xlr5-Xlr3-Xlr4* (Figure 0.6.3). The closest orthologs in human are the *FAM9* genes, located at the Xp22 region. Little is known about the function of these groups of genes. Within each paralog subfamily, e.g. *Xlr3a, Xlr3b*, and *Xlr3c*, nucleotide sequence conservation is nearly 98% and they code for identical proteins. The *Xlr* genes are closely related to *Synaptonemal complex protein 3 (Sycp3)*, which is involved in homologous chromosome pairing during meiosis. The Xlr3 protein harbors a conserved coiled-coil Cor1 domain, the function of which is still unclear. This Cor1

domain is a component of the chromosome core in the meiotic prophase chromosome [36]. The Cor1 domain is also found in Sycp3, and other meiotic proteins including Xlr, Slx/Smr, Slx2/Xlr6, and Sly–all of which are derived from *Sycp3* and are *Xlr* family members.

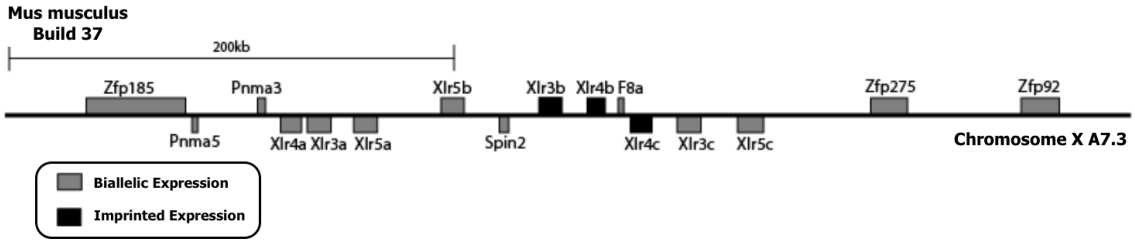


Figure 0.6.3: **Schematic representation of the *Xlr3/4/5* locus.** The paternally imprinted genes *Xlr3b, Xlr4b*, and *Xlr4c* sit in the middle of the locus with their biallelically expressed paralogs.

Along with genes related to neurodevelopment, the X-chromosome is also highly enriched for genes related to reproduction. Robert Foley previously showed that *Xlr3* becomes upregulated during spermatogenesis, and that the Xlr3 protein localizes to the XY sex body during male meiosis (Foley, unpublished). Ongoing work in the lab strives to understand the function of the Xlr3 protein with a novel knockdown transgenic mouse model.

0.7 Mechanisms of Epigenetic and Imprinting Regulation

To date, there have been over 150 imprinted genes identified in mouse and human, the vast majority of which are autosomal [37]. Imprinted genes tend to exist in clusters, which allows access to common regulatory elements, and in general, are controlled by an imprinting control region (ICR). ICRs are typically found by

identifying a differentially methylated region (DMR). For a DMR to be considered an ICR, it must be shown to directly affect gene expression, and be resistant to global demethylation following fertilization [26].

IGF2/H19 is among the most well-studied imprinted loci. This autosomal locus shows reciprocal expression and is regulated by a DMR upstream of *H19*, characterized by hypomethylation on the maternal allele, and hypermethylation on the paternal allele [38]. The zinc-finger protein CTCF binds to the hypomethylated maternal allele, allowing downstream enhancers to activate expression of maternal *H19*. The lack of insulator binding on the paternal allele, allows for a long-range interaction of downstream enhancers with the *IGF2* promoter.

0.7.1 Differential DNA Methylation

A long-standing question in this lab has been the search for a DMR within the *Xlr3/4* locus that controls imprinted expression of these genes. In 2001, Jackson-Grusby *et al.* used DNA microarray technology to compare *Dnmt1* knockout expression to wild type in mouse fibroblast cell lines [39]. Dnmt1 is a methyltransferase necessary for the maintenance of DNA methylation throughout the cell cycle. Imprinted genes were highly represented among those with a significant change in expression, with *Xlr3b* and *Rhox5* among those most highly upregulated. A few years later, Schulz *et al.* compared expression data between *Dmnt3l*$^{-/+}$ knockdown embryos to wild type [40]. A lack of *de novo* methylation during oogenesis resulted in aberrant expression of imprinted genes. Again, *Xlr3b* and *Rhox5* were among

those most highly upregulated. These studies suggest that DNA methylation plays a role in silencing these X-linked genes.

Seth Kasowitz previously performed a 5-aza-2-deoxycytidine (5-aza-dC) treatment on 39,X^M and 39,X^P mouse neonatal fibroblasts. 5-aza-dC is a demethylating agent that binds to DNA methyltransferases, preventing maintenance methylation from taking place. A time course experiment from 0 to 120 hours showed an increase in expression of *Xlr3b* on both alleles (Figure 0.7.4). However, while X^P expression increased nearly 200-fold over the course of the experiment, X^M showed an increase of only 2 - 3 fold. Relative expression between X^M and X^P has a nearly 300-fold difference in untreated samples, and this drops to a mere 3 - 6 fold difference with 5-aza-dC treatment. It is possible that this overall increase of *Xlr3b* expression on both alleles is simply consistent with DNA methylation's expected suppressive effect on gene expression. The decrease in relative expression, on the other hand, suggests that DNA methylation may be one of multiple epigenetic factors contributing to the imprinted gene expression of this locus.

0.7.2 Allele-Specific Histone Modification

The modification of the amino-terminal tails of the core histone proteins can alter the transcriptional activity of the DNA they associate with [3]. Modifications can include methylation, acetylation, phosphorylation, ubiquitination, sumoylation, and more. These modifications can impact gene expression by altering chromatin structure, or the way DNA wraps around histone proteins, thus changing a pro-

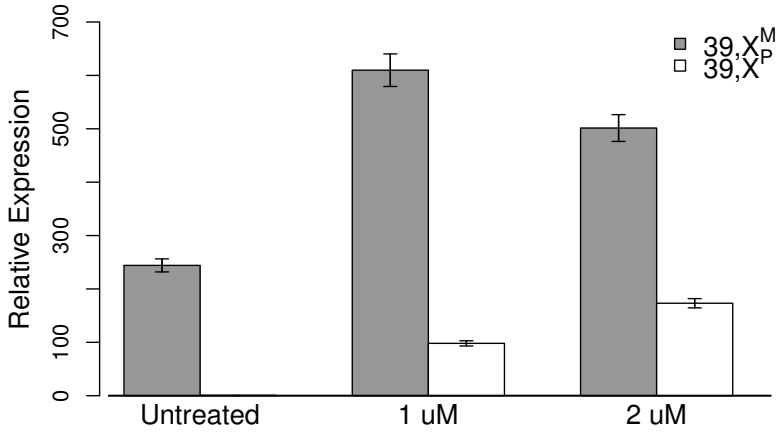


Figure 0.7.4: **Increased *Xlr3b* expression in 5-aza-2-deoxycytidine treated cell lines.** Real-Time PCR showing increased expression of *Xlr3b* in 39,X^M and 39,X^P cell lines, normalized to *ActB*. Error bars display 95% confidence. Courtesy of Seth Kasowitz.

moter's accessibility to transcriptional machinery. Allele-specific histone modifications (ASHM) have been found at several imprinted loci [41].

In 2006, Zupkovitz *et al.* compared expression data between HDAC1-deficient mouse embryonic stem (ES) cells to wild type [42]. HDAC1 is a histone deacetylase that represses transcription by catalyzing the removal of acetyl groups from core histone proteins. They identified several imprinted genes as novel targets of HDAC1. In HDAC1-deficient cells, *Xlr3b* was the gene that showed the greatest change, with a 54-fold increase in expression.

In another study, Fan *et al.* looked at the effect of linker histone H1 depletion in mouse ES cells [43]. They saw reduced local chromatin compaction and decreased global nucleosome spacing. Many genes whose expression was affected in triple-H1 null ES cells were imprinted or located on the X-chromosome. Again, *Xlr3b* was the most highly upregulated gene, with a 10.3-fold increase in expression. Reduced

DNA methylation within the regulatory regions of some of the H1 null affected genes suggests that linker histone H1 may help regulate imprinted gene expression by contributing to the maintenance of specific DNA methylation patterns, thus affecting expression of these genes.

Overall, it appears that more than one epigenetic mechanism is actively silencing *Xlr3b* and its paralogs. With regards to the imprinted expression, what still remains to be seen, is whether it is solely the paternal allele that is silenced by these mechanisms, or the maternal allele that is persistently expressed despite all attempts by the epigenome to silence it.

0.7.3 Interactome Analyses of Imprinted Domains

The aforementioned epigenetic mechanisms of regulation have profound effects in the form of driving chromatin loops within the nucleus. Chromatin loops are an essential component of gene regulation, as they allow for long-range contacts between genes and their regulatory elements. Chromatin models have suggested that imprinted genes are separated into active and inactive domains [44]. Chromosome conformation capture (3C) experiments have been used to show parent-specific looping interactions of *Igf2* and *H19* [45]. Another imprinted domain, *Dlk1/Gtl2*, has shown similar parent-specific interactions of the DMR, that only occur on the methylated paternal allele [46].

A *Hox9lacZ* transgene displayed imprinting when inserted into the *Integrin α-6 (Itga6)* gene, along with maternal-specific methylation, silencing, and the associated

histone modification profiles [47]. In addition, 3C experiments showed distinct nuclear architectures, with the maternal, repressed interaction profile showing a more closed and compact structure. Loss of *Dlx5* imprinting in MeCP2-null mice was associated with loss of silent chromatin loops at the *Dlx5-Dlx6* locus, which exist in wild type mice [48]. Instead, this region formed active chromatin-associated loops with other distant loci.

While the epigenome represents a level of gene regulation above the DNA sequence, the interactome forms an additional level in the distinct topology of chromatin that is determined by epigenetic mechanisms of gene regulation. Interactome analysis of imprinted domains can reveal distinct architecture between maternal and paternal genomes, reflecting the active and inactive status of the alleles. The chromosome conformation capture technique has proven to be a vital tool in the understanding of chromatin dynamics, organization, and differential regulation.

Chapter 1

Nuclear Organization of the Imprinted *Xlr3/4* Locus

1.1 Abstract

Using 4C-Seq, I characterized the interactome of the imprinted *Xlr3/4* locus, providing the first in-depth analysis of chromatin dynamics surrounding an imprinted gene cluster on the X-chromosome. I found significant differences in long-range contacts between X^M and X^P, with X^M showing far more genome-wide interactions than X^P. The 4C-identified regions show earlier DNA replication timing, enriched RNA pol II occupancy, and marks of active chromatin, suggesting these domains reside among open chromosomal compartments. Local interactions within a 2.7 Mb region surrounding the locus show the most dramatic differences between X^M and X^P, likely reflecting a mechanism that controls imprinted gene expression. Genes exclusively found in X^M domains are linked to behavioral abnormality, intellectual disability, cognitive impairment, as well as autism spectrum disorder. Taken together, these results identify *cis*-regulatory elements and reveal key differences in genome architecture of an imprinted X-linked cluster.

1.2 Background

A question of increasing interest in the field is how nuclear organization influences gene regulation. Advances in molecular techniques within the past decade and a half have allowed greater insight into the spatial organization of chromosomes, as well as the complex three-dimensional networks of chromosomal interactions that make up mammalian genomes. Studies of functional nuclear architecture reveal multiple levels of gene regulation, including but not limited to: long-range interactions between enhancers, repressors, and their target genes, the coordinated assembly of complex gene networks, and the critical role of epigenetic modifications in influencing chromatin state. Interactions in the nucleus can be dynamic, shifting or vanishing depending on cell type, developmental stage, and disease state. Studies of nuclear organization shed light on the functional consequences of these often-transient associations.

The chromosome territory model for nuclear organization proposes that there are regions of the nucleus that are preferentially occupied by particular chromosomes. This non-random organization of interphase chromosomes was first suggested by Carl Rabl in 1885, but the concept wasn't popularized until nearly a century later when light microscopy and fluorescent *in situ* hybridization (FISH) techniques allowed for the direct visualization of individual chromosome territories [49]. Chromosome territories facilitate the communication between genes and their regulatory elements, as the location of a gene within a chromosome territory can influence its

access to machinery for specific nuclear functions, such as transcription, DNA repair, and splicing [50]. Chromatin looping in the nucleus allows genomic elements widely spaced in the linear genome to come into close spatial proximity, and can allow for the co-localization of genes at subnuclear structures, altogether forming a topological model for gene regulation.

The development of chromosome conformation capture (3C) technology and its genomic variants has paved the way for analysis of nuclear organization at an unprecedented (and rapidly increasing) resolution and throughput. The 3C methodology relies on formaldehyde fixation of chromatin, followed by digestion and re-ligation of DNA fragments, capturing sequences that are in close proximity at the time of fixation. Chimeric ligation products can be quantified by various means, from qRT-PCR to deep sequencing, creating maps of DNA contact frequencies and gaining insight into chromosome topology inside the cell. These DNA contact maps are only the first step in determining how genome shape and structure relates to genome function. A key question to keep in mind: does genome organization affect gene function or is it a reflection of it? It is also important to note that 3C techniques provide an averaged conformation of the genome, usually from hundreds of thousands, sometimes millions of cell nuclei, thus creating a "consensus snapshot" of the nucleus, which is highly dynamic and highly variable among individuals cells. With this in mind, 3C techniques can provide a plethora of information about the dynamic functional states of the genome.

Figure 1.2.1 provides an overview of a few 3C-derived methods. Briefly, 3C-

qPCR analyzes one ligation junction at a time. 4C utilizes a secondary digestion and ligation step to create a linear profile of all contacts of a particular bait fragment. 5C and Hi-C create two-dimensional interaction heat maps of a particular region and genome-wide, respectively. ChIP-loop and ChIA-PET combine 3C with chromatin immunoprecipitation (ChIP) to enrich for interactions bound by specific proteins of interest. Several other 3C techniques have been developed, each with their own resolution, advantages, and pitfalls.

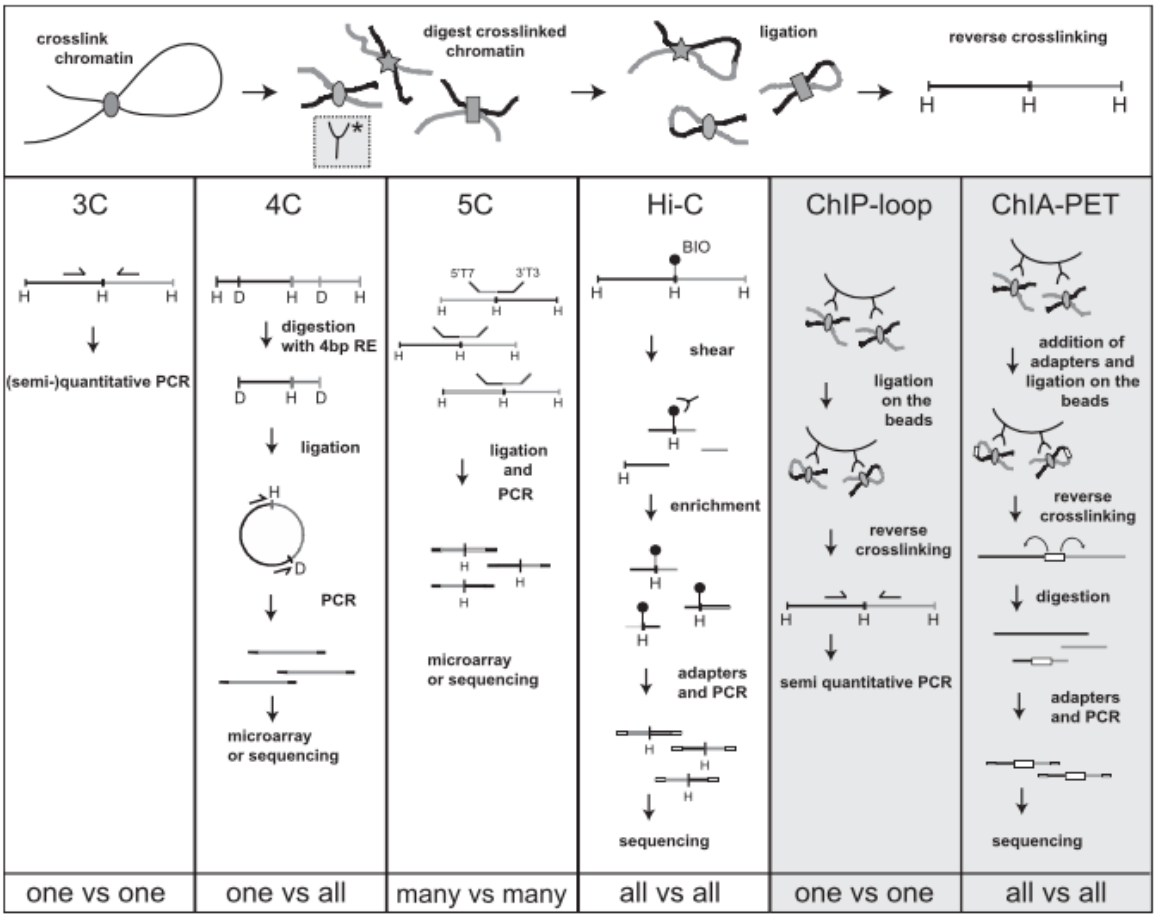


Figure 1.2.1: **Overview of 3C-derived methods.** Formaldehyde crosslinking, restriction digestion, and ligation steps are common to all 3C-derived methods. Adapted from de Wit *et al.* [51].

A key characteristic of all 3C-related techniques is the high capture probability between neighboring restriction fragments. As the very nature of these methods is to capture regions in close spatial proximity, by their linear location, neighboring fragments must be overrepresented in genomic analyses. Contact probability decreases with increasing genomic distance until it reaches a baseline level, reflecting background interaction frequencies of fragments randomly contacting in the nucleus [52]. This rapid decline in contact probability suggests that any two fragments that interact greater than expected might represent a *bona fide* genomic interaction. For these types of higher-order datasets, windowed approaches are necessary for the analysis of long-range interactions, as they describe a statistical interpretation of two regions that have a higher probability of making contact when compared to other regions at a similar distance.

The circular chromosome conformation capture (4C) technique provides a high-resolution analysis of a specific locus and all of its genome-wide contacts. 4C is able to capture both intra- and inter-chromosomal long-range interactions and has been classically used with the aim of defining *cis*-regulatory elements controlling a single gene or gene cluster. Simonis *et al.* used 4C to show how the active β-globin locus in fetal liver tends to interact with actively transcribed genes, whereas the same inactive locus in fetal brain contacts transcriptionally silent loci [53]. This work demonstrates not only differences in contacts, but that chromosomes are able to fold into areas of active and inactive chromatin. Imprinted loci have also demonstrated unique parent-specific chromatin folding. 3C-qPCR and 4C assays have shown

that maternal and paternal alleles can associate with different elements, reflecting epigenetically regulated chromosomal networks [44][54].

Actively transcribed genes co-localizing and forming clusters of active domains with intervening inactive chromatin looping out has been termed active chromatin hubs (ACHs) [55]. These long-range interactions frequently involve a number of regulatory DNA elements. DNA hypersensitivity sites are hallmarks of gene regulatory elements, and often mark regions of the genome prone to high interaction profiles. Boundaries of active and inactive domains, termed topologically associating domains (TADs), are marked with CTCF and cohesin, complexes known to mediate DNA looping [56]. Recent studies have shown that binding polarity determines chromatin looping, which preferentially forms between convergent-oriented CTCF binding sites [57][58]. While cohesin association persists, inverted CTCF sites fail to form loops, and this altered chromatin landscape can lead to changes in gene expression. CTCF depletion results in a dramatic loss of TAD insulation, but does not affect higher-order chromosome folding, defining an important role for CTCF in creating focal boundaries in chromosome organization [59].

The three-dimensional nuclear architecture is pivotal in the regulation of transcription and in the maintenance of epigenetic states. Alterations in chromatin structure have been observed in aberrant conditions such as cancer, disease state, and loss of imprinting (LOI) [60]. Imprinted genes on the X-chromosome have been implicated in cognitive function and social impairment, as seen in patients with Turner syndrome and autism spectrum disorder (ASD). By combining chromosome

conformation capture with high-throughput sequencing, I examined long-range interactions and the nuclear architecture of an X-linked imprinted locus.

I sought to explore the global differences of interaction profiles of imprinted genes *Xlr3b* and *Xlr4c* in neonatal mouse brain, where there is greater than 50-fold difference in gene expression between maternal and paternal alleles. My data shows all interacting partners of our locus of interest and suggests overall differences between X^M and X^P, with the most dramatic differences occurring in a 2.7 Mb region surrounding the imprinted locus. The contact landscape revealed from this study sheds light on *cis*-regulatory elements, as well as networks relating to cognitive function and impairment.

1.3 Results

1.3.1 Generating a library of genome-wide chromatin interactions

To characterize the interactome of the imprinted *Xlr3/4* locus, 4C was used, coupled with next-generation sequencing. This 4C-Seq strategy is used to uncover all genomic loci that are interacting with a "bait" locus, specified by the design of inverse PCR primers. In brief, chromatin was crosslinked with formaldehyde, followed by enzymatic digestion with either HindIII (ˆAAGCTT) or DpnII (ˆGATC), and ligation to capture interacting sequences. A secondary digestion and ligation circularized the template, allowing for the amplification of all bait-ligated interactor sequences by inverse PCR. The resulting library was sequenced by high-throughput methods.

Three replicate libraries for X^M and X^P mouse neonatal cortex (P0) were made

using either HindIII or DpnII as the primary restriction enzyme. The size of the restriction fragments generated by the enzyme of choice determines the resolution of interaction mapping. A 6 bp-cutter such as HindIII is a traditional enzyme of choice. However, a 4 bp-cutter such as DpnII provides higher resolution and can be used to measure more specific interactions of a given region. Inverse PCR primers were designed in *Xlr3b* and *Xlr4c* (Figure 1.3.1). Two types of libraries were generated for each gene with three biological replicates per library.

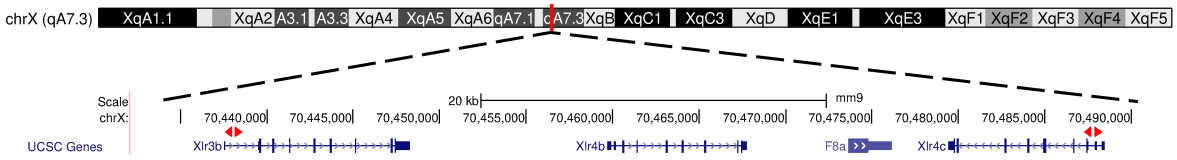


Figure 1.3.1: **Overview of 4C-Seq bait region at the imprinted *Xlr3/4* locus.** A detailed view of the *Xlr3/4* locus at the qA7.3 region of the mouse X-chromosome. Primers for inverse PCR shown by red triangles.

4C libraries were analyzed with 4Cker, a pipeline that maps reads to a reduced genome of unique sequences directly adjacent to all primary restriction enzyme sites in the genome. Libraries were evaluated for reproducibility and validity. The majority achieved an overall *cis* ratio >40%, a criteria generally accepted to be indicative of good quality experiments [61]. Of particular interest, X^M libraries consistently showed more reads in *cis* compared to X^P libraries (Figure 1.3.2).

1.3.2 Long-range interactions at the imprinted *Xlr3/4* locus

A Hidden-Markov Model based approach was applied to identify significant domains that interact with each 4C bait locus. Differential interactions were deter-

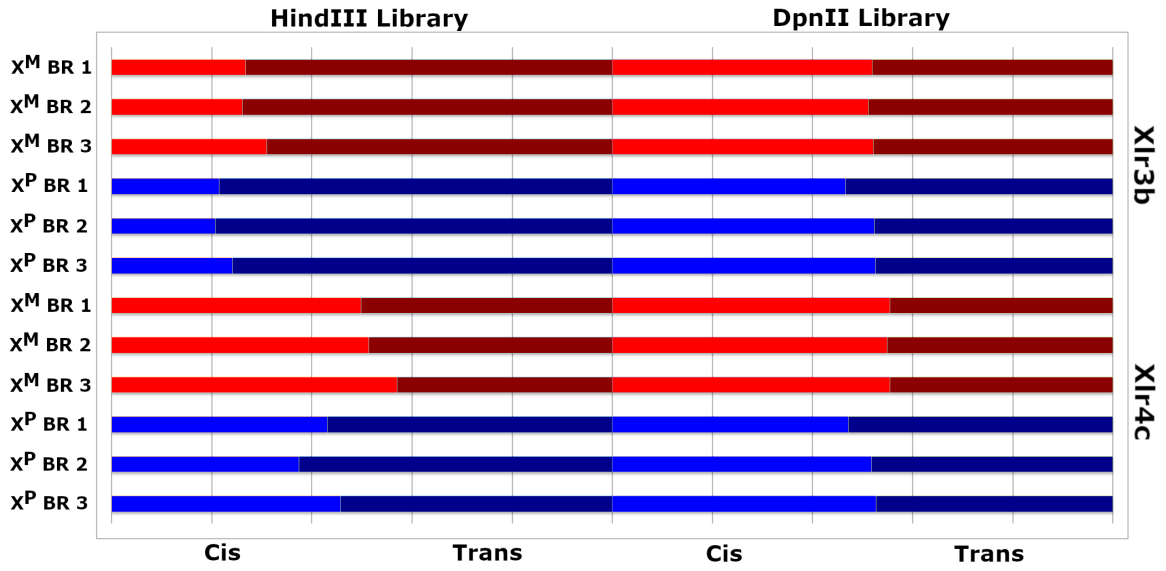


Figure 1.3.2: **Distribution of 4C-Seq reads.** Stacked bar plot showing the percentage of mapped reads in *cis* and in *trans* in *Xlr3b* and *Xlr4c* libraries. Three X^M (red) and X^P (blue) biological replicates are included. See Table B.1.9 for number of mapped reads.

mined for X^M and X^P with biological replicates. Adaptive window sizes were used to correct for differences in signal coverage based on the distance from the bait, to calculate significant interactions in near-bait regions, typically 2 - 10 Mb surrounding the bait, and on far-*cis* and *trans* chromosomes. The number and distribution of high-fidelity interacting sites are shown in Figure 1.3.3.

Of the HindIII libraries generated, *Xlr3b* showed few interactions on the X-chromosome and many more X^M interactions in *trans* than X^P (Figure 1.3.3A). *Xlr4c* HindIII libraries found high interacting domains across much of the genome and likely reflect unreliable data due to promiscuous primer binding and poor primer design (data not shown). Of the DpnII libraries generated, both *Xlr3b* and *Xlr4c* produced results comparable to similar studies across the X-chromosome (Figure

1.3.3B). However, *trans* interacting domains created using the 4 bp-cutter restriction enzyme called too many domains and again, likely reflect limitations of the 4C technique (data not shown).

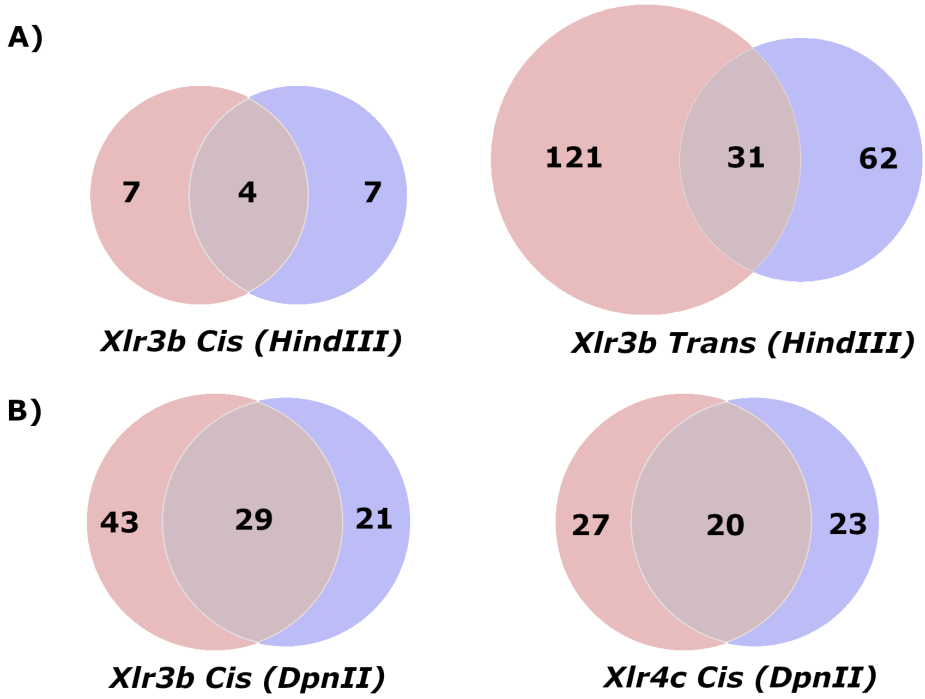


Figure 1.3.3: **Distribution of 4C interactions.** Venn diagram showing the number of overlapping significant interactions between X^M (red) and X^P (blue). A) Libraries fragmented with HindIII with interacting domains shown in *cis* (k = 10) and in *trans* (k = 18) using *Xlr3b* as bait. B) Libraries fragmented with DpnII with interacting domains in *cis* using *Xlr3b* (k = 15) and *Xlr4c* (k = 15) as bait.

For almost every window size and library analyzed, X^M showed a greater number of significant interactions compared to X^P. Significant interacting domains of *Xlr3b* (HindIII) in *cis* and in *trans* are shown in Figures 1.3.4A and 1.3.4B. There were few long range intrachromosomal interactions with the majority directly surrounding the bait locus, and many more interchromosomal interactions. The interchromosomal profile shows X^M making more *trans* contacts compared to X^P, suggesting that the

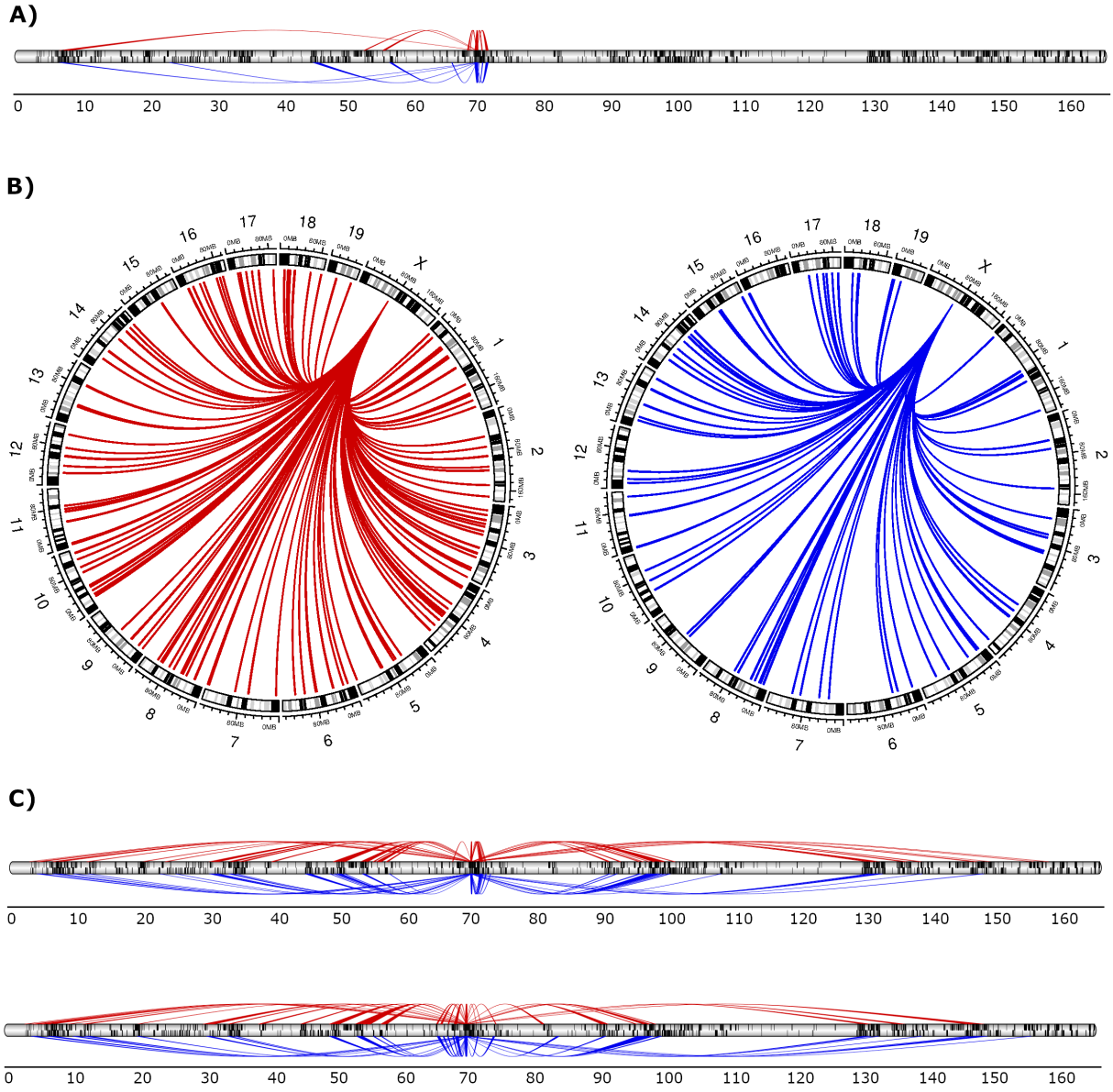


Figure 1.3.4: **Significant 4C domains in *cis* and in *trans*.** Distribution of significant domains in X^M (red) and X^P (blue). A) Spider plot showing long-range interactions in *cis* across the X-chromosome with *Xlr3b* as bait (HindIII). B) Circos plot showing long-range inter-chromosomal *trans* interactions with *Xlr3b* as bait (HindIII). C) Spider plots showing long-range interactions in *cis* across the X-chromosome with *Xlr3b* (top) and *Xlr4c* (bottom) as baits (DpnII).

chromosome organization around *Xlr3b* is very different. These results indicate that the maternal and paternal loci may be localizing to different parts of the cell nucleus, thus one making more physical contacts with other chromosomes than the other.

Significant interacting domains of *Xlr3b* and *Xlr4c* (DpnII) across the X-chromosome are shown in Figure 1.3.4C. The spider plots show the similarities of long-range interactions in *cis* between X^M and X^P. This is not necessarily surprising, as the intra-chromosomal profile of imprinted loci tend to be quite similar outside the locus's TAD, with the majority of differences occurring within it [62]. This will be further addressed in Section 1.3.4.

The distribution of 4C domains unique to either X^M or X^P across the X-chromosome is shown in Figure 1.3.5. This figure also shows 4C domains that were found in both X^M and X^P libraries. While there is significant overlap between X^M and X^P in all datasets, there are also many significant interactions only found in one or the other.

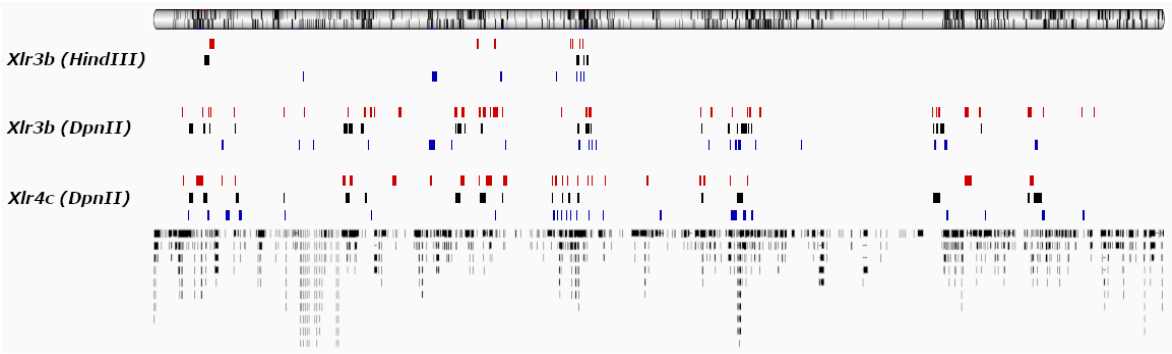


Figure 1.3.5: **Long-range *cis* interactions of *Xlr3b* and *Xlr4c*.** Distribution of significant domains across the X-chromosome, represented by colored bars (interactions unique to X^M, red; interactions unique to X^P, blue; interactions found in both, black) shown above gene density.

1.3.3 4C domains reside in open chromosomal compartments

After identifying all highly interacting regions, I set out to characterize the nuclear properties of these 4C domains. A widely accepted model popularized by Hi-C studies suggests that chromosomes are organized into open and closed compartments [52]. I speculated that the 4C domains would show characteristics of open chromosomal compartments. Earlier DNA replication timing values of the 4C domains support this hypothesis.

Figure 1.3.6 shows the distribution of DNA replication timing values of *Xlr4c* X^M and X^P 4C domains (DpnII) compared to the whole genome in ten pluripotent cell lines. These regions display earlier replication timing in every cell line shown. Replication timing is correlated with the expression of genes, with open, accessible regions being the first to replicate. Similar to the two-compartment model proposed by Hi-C studies, replication timing values also suggest a model in which early-replicating and late-replicating regions are organized in a segregated manner within the cell nucleus [63]. Early replication takes place in the interior of the nucleus, with later replication occurring near the periphery, further facilitating a non-random model of nuclear organization.

Another characteristic of the 4C domains is that they are enriched for RNA pol II occupancy and marks of open, active chromatin [64]. Histone modifications are involved in the control of transcription. For actively transcribed genes, H3K4me3 is often found at the promoter, while H3K36me3 is deposited along the active gene

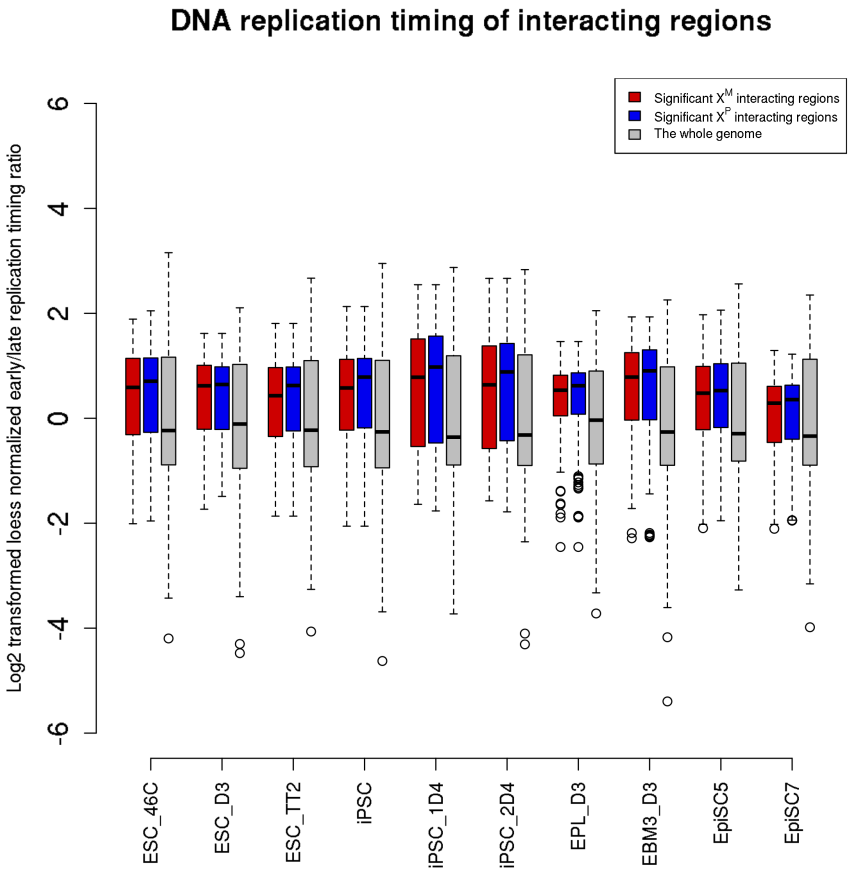


Figure 1.3.6: **DNA replication timing of 4C domains.** Box plot showing the distribution of DNA replication timing values of *Xlr4c* 4C interacting regions in X^M (red) and X^P (blue) compared to the whole genome in 10 pluripotent cell lines. Early replication domains have the logarithm of replication timing ratio > 0.

body [65]. I examined the enrichment of the aforementioned histone marks and found H3K4me3 and H3K36me3 were enriched around *Xlr4c* 4C sites (+/- 500 kb), in both X^M and X^P (Figure 1.3.7). These hallmarks of open chromatin suggest 4C sites tend to be located near actively transcribed genes. Enriched RNA pol II occupancy further supports a strong association with regions actively poised for transcription. Surprisingly, I also found enrichment of H3K27me3 in both X^M and X^P, a modification

associated with repressed genes and facultative heterochromatin. H3K27me3 acts in opposition to H3K4me3, thus their mutual enrichments were unexpected.

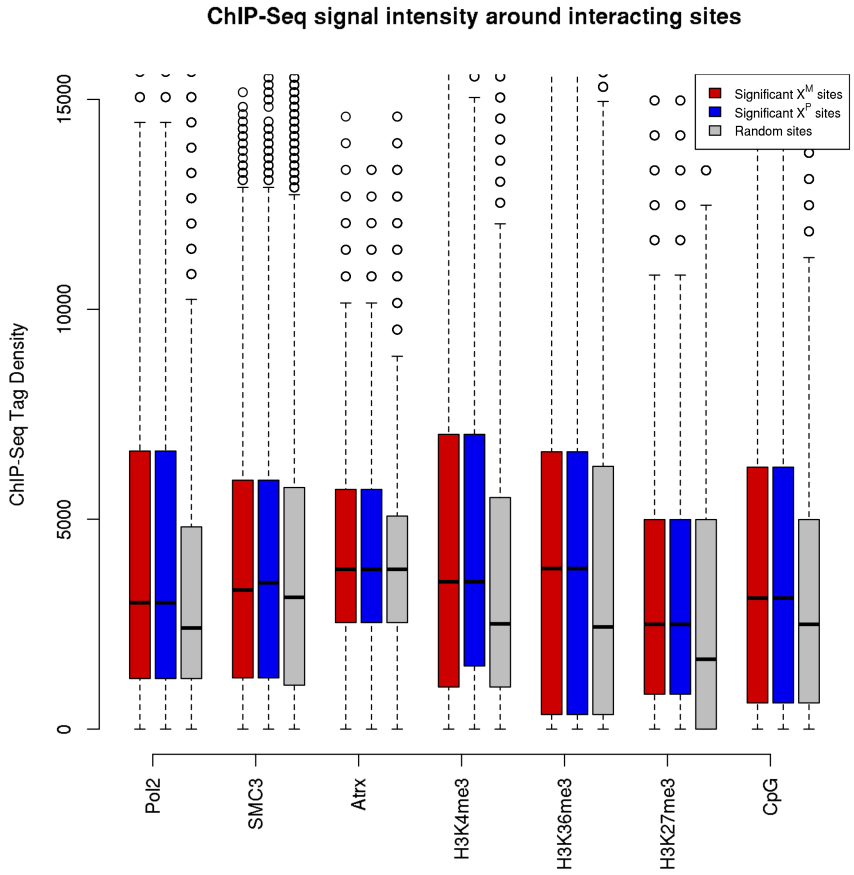


Figure 1.3.7: **Enrichment of active chromatin signatures around 4C domains.** Boxplot of different chromatin marks and transcription factor binding distribution around the *Xlr4c* interacting sites in X^M (red) and X^P (blue) compared to random sites (grey). ChIP-seq tags within +/- 500 kb around interacting sites were counted and normalized to 10 million.

SMC3, a protein that is part of the cohesin complex, and Atrx, a transcriptional regulator, were both marginally enriched around 4C sites. The cohesin complex plays a role in genome organization, and is frequently found in areas of active transcription, where it co-localizes with activator complexes [66]. Atrx, a chromatin

remodeler, works in conjunction with cohesin to regulate the postnatal expression of certain imprinted genes in the brain [67]. Atrx is also known to bind to regions associated with CpG islands, another genomic feature found to be enriched around 4C sites [68].

Additionally, annotations of interspersed repeats and repetitive elements within *Xlr4c cis*-interacting domains show an enrichment of SINE repeats and a depletion of LINE repeats, which suggests that these domains reside in gene-dense regions of the chromosome (Figure 1.3.8). Compared to the rest of the genome, the X-chromosome is distinct in its enrichment of LINE elements. Described as the Lyon repeat hypothesis, this distinction is proposed to favor X-chromosome inactivation [69]. The nearly two-fold enrichment of LINEs makes the X fundamentally distinct from autosomes and helps spread the inactivation signal across the X-chromosome during XCI. Within the 4C domains of the X-chromosome, there is still an enrichment of LINEs compared to the whole genome. But compared to the X-chromosome as a whole, there is a depletion of these repetitive elements, presumably because interactions are more likely to occur within functional regions of the genome.

1.3.4 Short-range comprehensive analysis of looping interactions

Very little is known about the regulatory elements that control the imprinted expression of the *Xlr3/4* locus. As previously mentioned, the greatest differences in interactions at imprinted loci tend to exist within the locus's TAD. TADs represent DNA regions that are enriched in chromatin contacts [70]. Loci located in different

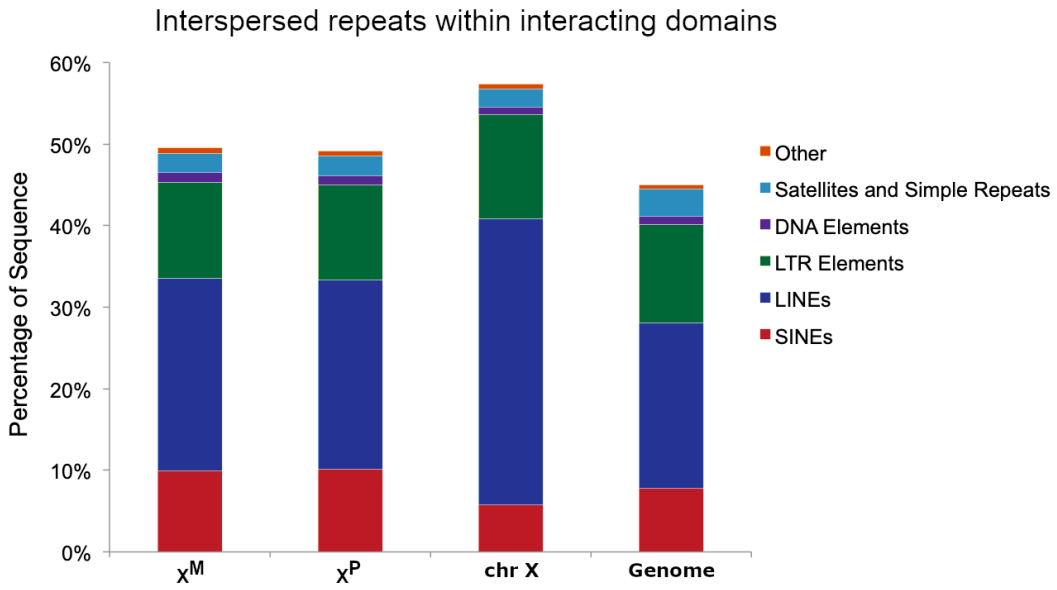


Figure 1.3.8: **Repetitive elements within 4C domains.** Annotation of interspersed repeats and low complexity DNA sequences in *Xlr4c* 4C domains in X^M and X^P compared to the X-chromosome and whole genome.

TADs interact less frequently than those within the same TAD. Short-range comprehensive analysis of looping interactions around the *Xlr3/4* locus will help shed light on regulatory elements that could be involved in the control of imprinted gene expression.

Differential interactions within a region +/- 2 Mb of the imprinted *Xlr3/4* locus are shown in Figure 1.3.9. 4C domains for *Xlr3b* and *Xlr4c* (DpnII) libraries display many differences between X^M and X^P datasets. *Xlr4c* domains show a greater number interactions and more drastic differences in the region analyzed compared to *Xlr3b*. This may be due to differences in primer efficiency or genuine interaction differences between the two genes.

To better analyze the region where these interactions occur, I used Hi-C data

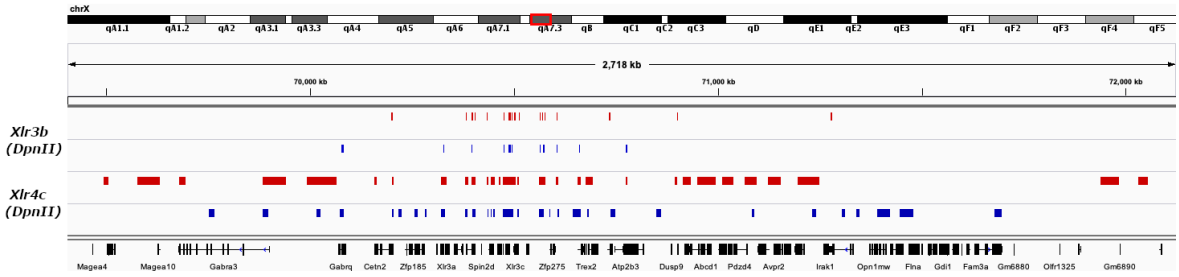


Figure 1.3.9: **Near-bait interactions of *Xlr3b* and *Xlr4c*.** Distribution of significant domains within +/- 2 Mb surrounding *Xlr3b* (k = 5) and *Xlr4c* (k = 5) in X^M (red) and X^P (blue).

generated from a CH12 cell line. Figure 1.3.10A shows the TAD that contains the imprinted *Xlr3/4* locus and is the region where the most interactions are expected to occur. My *Xlr4c* 4C data, which shows significant differences between X^M and X^P in a 2.7 Mb region, is located within a sub-TAD (Figure 1.3.10B). These interaction differences are likely related to the regulation of *Xlr3/4* imprinted gene expression.

To ensure that the 4C data is accurate and reliable, I used results of a single primer from a chromosome conformation capture carbon copy (5C) assay to verify novel contacts near the locus. X^M and X^P 5C libraries were made with monosomic mouse primary fibroblasts, using NcoI (ˆCCATGG) as the primary restriction enzyme. 5C technology relies on ligation-mediated amplification (LMA), where forward and reverse 5C primers are annealed and ligated at 3C junctions in a multiplexed manner. Figure 1.3.11 shows the alternating 5C primer scheme used in this experiment.

Using a 5C primer designed in *Xlr3b* as an anchor (Xlr3b-FOR-81), the linear interaction profile of X^M and X^P within a 465 kb region is shown in Figure 1.3.12. When aligned with *Xlr4c* 4C interacting domains, I observed high similarity between the interaction profiles, despite the anchors being in different genes. Due to the near

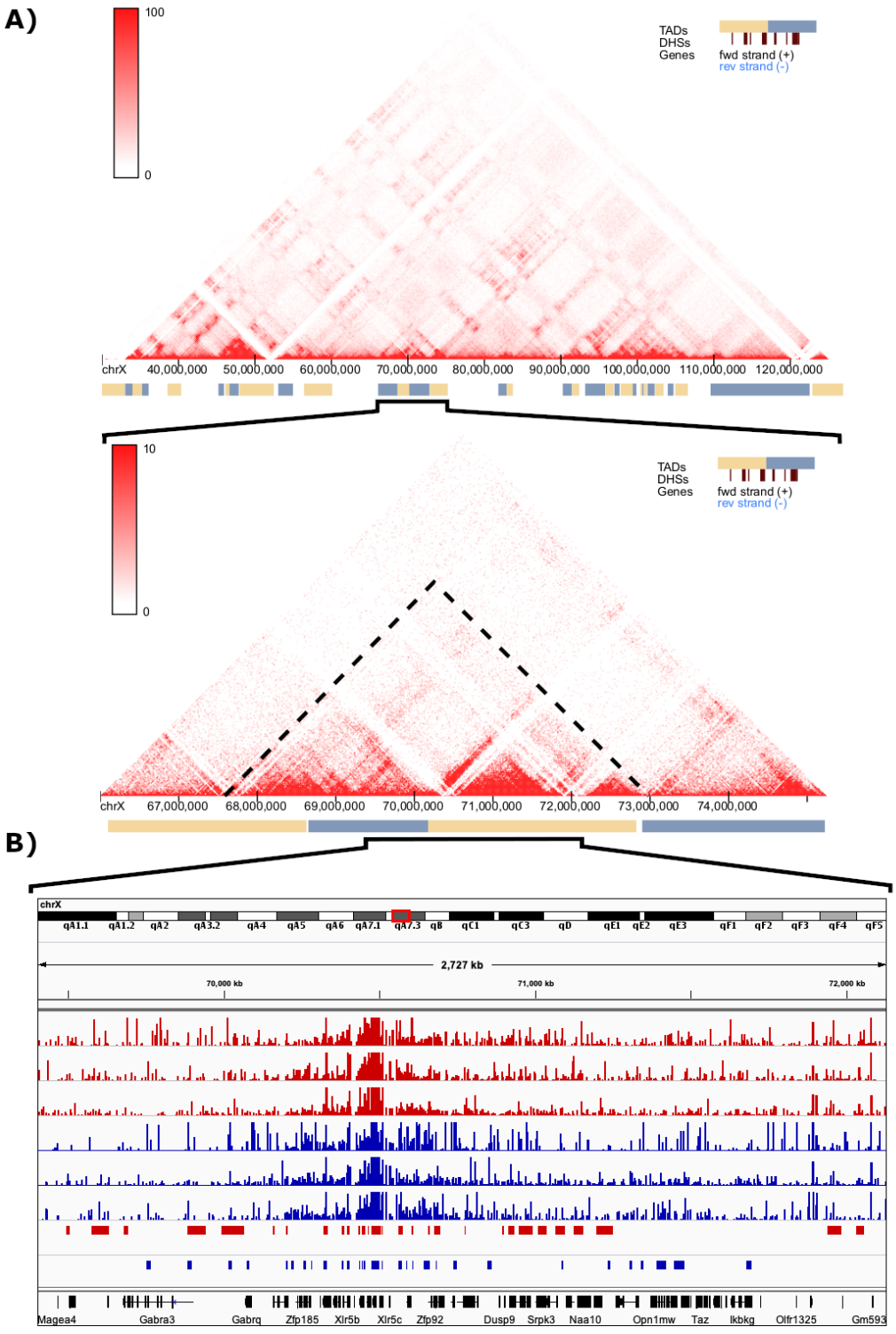


Figure 1.3.10: **Near-bait interactions show parent-specific differences within a TAD.** A) Hi-C data from CH12 cell line showing TAD organization. The *Xlr3/4* locus is further contained within a sub-TAD, marked by dashed lines, shown at 25 kb resolution. Data from Rao *et al.* [71]. B) Raw mapped reads and significant *Xlr4c* 4C interactions in X^M (red) and X^P (blue) of the region marked in the heatmap.

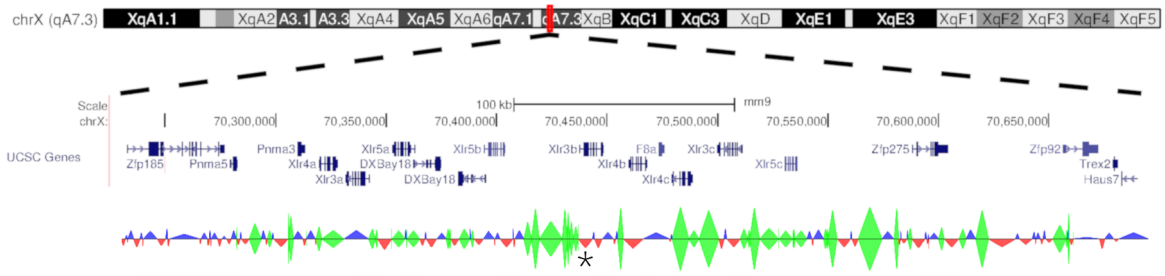


Figure 1.3.11: **Overview of 5C-Seq primer design at the imprinted *Xlr3/4* locus.** A detailed view of the *Xlr3/4* locus at the qA7.3 region of the mouse X-chromosome with an alternating design scheme (forward primer, red; reverse primer, blue; no primer, green). The height of each triangle scales with the repetitiveness of the primer at the specific genomic location. Primer Xlr3b-FOR-81 denoted with asterisk.

identical sequences of *Xlr4b* and *Xlr4c*, no 5C primers were made in *Xlr4c* and I could not compare the gene's 5C interaction profile with the 4C dataset. Despite this, there were many similarities observed between the *Xlr3b* 5C interaction profile and the *Xlr4c* 4C interacting domains. Differences between them can be attributed to the anchors being in different genes, as well as different restriction enzymes used in the experiments, thereby affecting resolution, and inherent differences and limitations of each technique.

Overall, the 5C-Seq technique did not yield reliable results, due to the paralogous nature of the region of interest. However, the linear profile from a single 5C primer in *Xlr3b* suggests that the 4C analysis is reliable, and that the long-range contacts uncovered represent putative regulatory elements, which could be relevant in the regulation of this imprinted locus.

Next, I set out to see if these contacting regions display features of active enhancers using Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) and CTCF, p300 ChIP-Seq datasets. ATAC-Seq aims to identify ac-

cessible regions of DNA and is considered a newer alternative to DNase-Seq. Hypersensitivity to DNase I is a common attribute of functional elements, as these nucleosome-free, open and accessible regions of the genome are presumably bound by regulatory factors. ATAC-Seq follows this same principle.

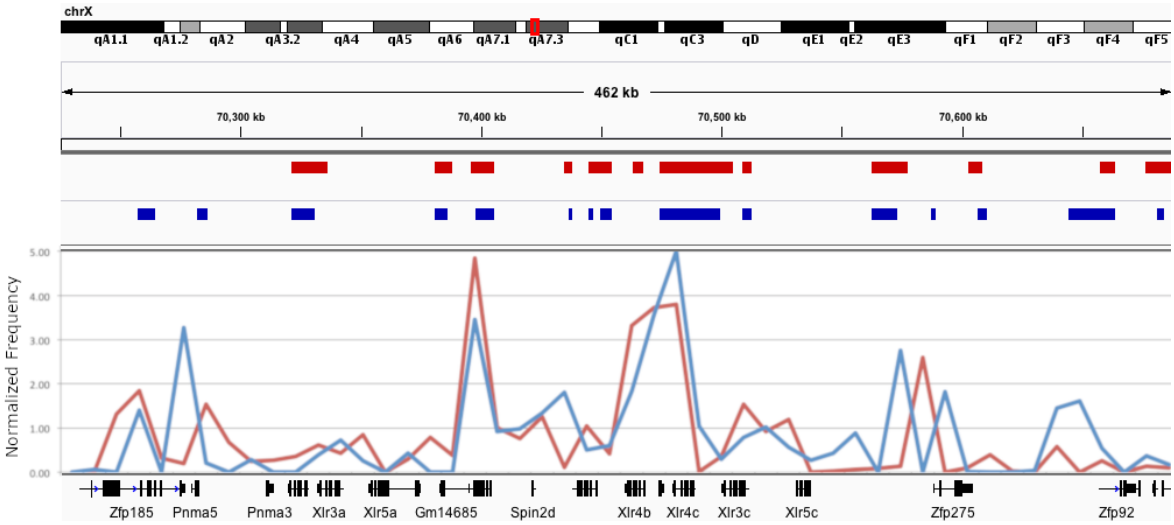


Figure 1.3.12: **5C interaction profile of *Xlr3b*.** *Xlr3b* 5C interaction profile (Xlr3b-FOR-81, Table A.1.4) aligned with *Xlr4c* 4C significant domains in X^M (red) and X^P (blue).

The distribution of ATAC-Seq peaks correlates well with the highest density interacting regions of the Hi-C heatmap (Figure 1.3.13). In this figure, sub-TAD boundaries overlap with CTCF peaks and are marked by blue triangles. A large majority of ATAC-Seq peaks are contained within these boundaries. Furthermore, many regions of chromatin interactions coincide with these ATAC-Seq peaks, in agreement with the previous notion that interactions occur at or near open chromosomal compartments.

Many of the *Xlr4c* 4C domains overlap with two hallmarks of DNA regulatory

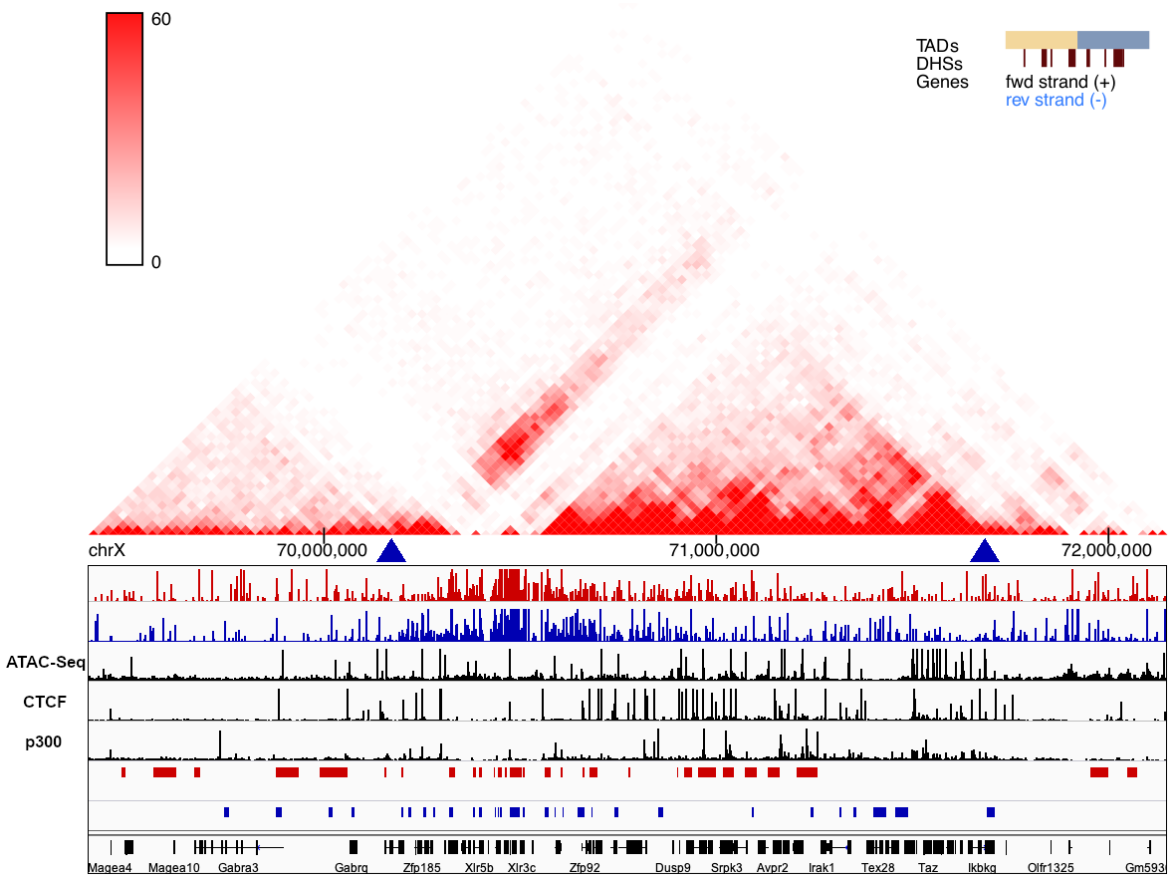


Figure 1.3.13: **4C interacting regions display characteristics of regulatory elements.** Hi-C heatmap and *Xlr4c* 4C interaction profile aligned with ATAC-Seq, CTCF and p300 ChIP-Seq profiles over a 2.7 Mb region chracterized. Boundaries of sub-TAD marked by blue triangles. 4C profile and significant domains represent X^M (red) and X^P (blue). ATAC-Seq data courtesy of Glenn Milton. ChIP-Seq data from Rao *et al.* [71].

elements: CTCF and p300. CTCF functions as a genome organizer and dimerizes to join clusters of genes with co-ordinated expression or bring together genes and their regulatory elements [56]. CTCF is well known to correlate with looping interactions and can be considered an enhancer facilitator. p300 is a mark of active enhancers. Several prominent interactions also coincide with p300 peaks, altogether suggesting that the three-dimensional organization characterized using 4C-Seq brings long-

range regulatory elements into close proximity with the imprinted *Xlr3/4* genes.

The publicly available datasets used here have several limitations. The Hi-C and ChIP-Seq datasets used CH12 cell lines, which are female, and would contain both X^M and X^P cells. Any interactions or signal seen here could be attributed to either or both chromosomes. Additionally, due to the paralogous nature of these genes of interest, there has always been an inherent difficulty in mapping to this region of the X-chromosome. A lack of signal in a region could be a mapping artifact as opposed to a *bona fide* absence of a biological occurrence. This is particularly evident in the Hi-C heatmap, in a region almost completely free of interaction signal at about 70.4 Mb, which represents our imprinted locus. With the advent of longer read sequencing, it is hopeful that these limitations will soon be overcome.

1.3.5 4C domains enriched with genes linked to cognitive impairment

4C interacting domains display a strong association with chromatin marks characteristic of active transcription and increased RNA pol II occupancy. As the *Xlr3/4* genes have been implicated in cognitive impairment and ASD, I hypothesized that the 4C domains might identify other genes involved in neurocognition and neurodevelopment. Nearly half of all interacting regions are located within 50 kb of a gene. After identifying regions that are unique to either X^M or X^P 4C libraries, and regions that are present in both, I used Genomic Regions Enrichment of Annotations Tool (GREAT) to analyze the functional significance of these interacting sequences.

GREAT determines statistical enrichment by associating genomic regions with

nearby genes and then applying gene annotations to those regions [72]. Every gene is assigned a basal regulatory domain of 5 kb upstream and 1 kb downstream of the TSS, extended in both directions to the nearest gene's basal domain up to 1000 kb in one direction. Each genomic region used as input, in this case 4C interacting domains, is associated with all genes whose regulatory domain it overlaps. Several ontologies were used to apply functional significance to interacting regions. The Human Phenotype Ontology has 6,096 terms covering 2,746 (13%) of all 21,176 genes, and 239,209 term-gene associations. The Disease Ontology has 2,209 terms covering 7,519 (35%) of all 21,176 genes, and 224,940 term-gene associations.

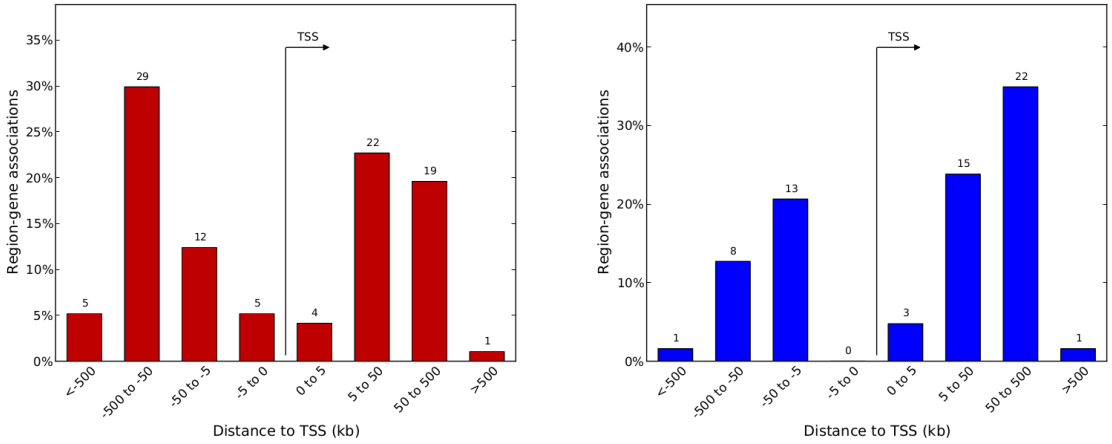


Figure 1.3.14: **Distribution of region-gene associations within 4C domains.** Distribution of *Xlr4c* 4C domains binned by orientation and distance to TSS in X^M (red) and X^P (blue).

Figure 1.3.14 shows the distribution of *Xlr4c* interacting regions binned by orientation and distance to TSS, with domains only found in X^M showing a slight preference for upstream of the TSS and domains only found in X^P showing a slight preference for downstream of the TSS. Among the *cis*-interacting regions of *Xlr4c*,

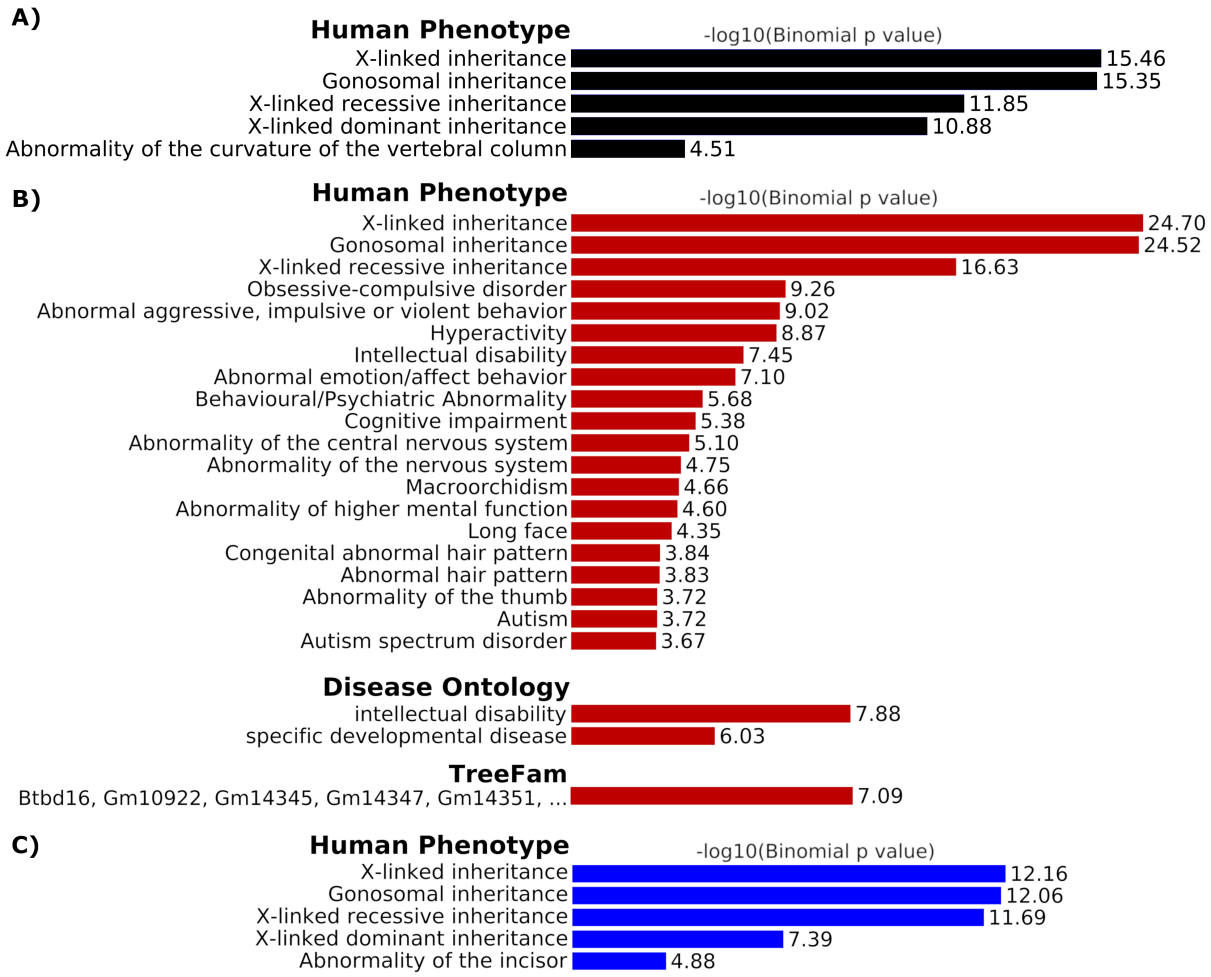


Figure 1.3.15: **Functional annotation of genes within *Xlr4c* 4C domains.** Prediction of functions of *cis*-regulatory regions common to X^M and X^P (A), found exclusively in X^M (B), and found exclusively in X^P (C).

I observed significant enrichment of genes related to behavioral abnormality, intellectual disability, and cognitive impairment. This is not unexpected, as the X-chromosome is enriched for genes related to intelligence and cognition. What is surprising, is that this enrichment was found only in regions unique to X^M (Figure 1.3.15, Table 1.3.1). In addition, genes related to autism and autism spectrum disorder (ASD) are also enriched in X^M domains. By contrast, four out of five hits in X^P

domains and regions found in both X^M and X^P are related to X-linked inheritance or gonosomal inheritance.

Table 1.3.1: Unique *cis* interactions of *Xlr4c*

Library	Genes
39, X^M	1700020N15Rik, 4930408F14Rik, 4930567H17Rik, 4933436I01Rik, *Aff2*, Araf, *Arhgef9*, Atp11c, *Atp6ap2*, Awat1, BC023829, Bmp15, Cd99l2, Cdr1, Chic1, Ctag2, Cxx1c, Dgat2l6, Etd, *Fgd1*, *Fmr1*, Gm10922, Gm14345, Gm14862, Gm20489, Gm2790, Gm2799, Gm2913, Gm2927, Gm5640, Gm6890, Gpr64, *Gria3*, Hmgb3, *Huwe1*, Idh3g, *Iqsec2*, Klf8, Lancl3, *Magt1*, Mamld1, Map3k15, *Med12*, Mid2, Mtmr1, Ndufb11, Nudt10, *Pcdh19*, *Phf8*, *Phka2*, Prkx, *Rab39b*, Rgn, Rps4x, Rragb, Sh3kbp1, Slitrk4, *Smc1a*, *Sox3*, Spin4, Srpx, Ssr4, Stard8, Sytl5, Tbl1x, Thoc2, Timp1, Tsr2, Tsx, Vbp1, Vgll1, Yipf6, Zic3
39, X^P	4930480E11Rik, 4930525M21Rik, AV320801, Apoo, Atp2b3, Bgn, Bhlhb9, Ccnb3, Cenpi, Chrdl1, Cnga2, Ddx3x, Dgkk, Dmrtc1b, Dmrtc1c1, Eda, Fgf16, Gab3, Gata1, Glod5, Gm14698, Gm4906, Gpc3, Gpc4, Gpr119, Gpr173, Gprasp2, Ikbkg, Magea4, Maged1, Mcf2, Nhs, Nsdhl, Ogt, Olfr1326-ps1, Pak3, Pramel3, Prrg1, Rai2, Rhox9, Slc25a14, Taf1, Tbl1x, Tmem28, Tmem35, Tmsb15b1, Tspyl2, Usp9x, Vma21, Zcchc18, Zfp185

**Italicized genes implicated in intellectual disability, cognitive impairment, and/or ASD [73].*

GREAT analyzes the functional significance of *cis*-regulatory regions by modeling the genome regulatory landscape and computing enrichment using a binomial test that accounts for variability in domain size by measuring the total fraction of the genome annotated for any given ontology term and counting how many input domains fall into those regions [72]. For these analyses, statistical enrichment is

evaluated relative to a background set of the entire genome. So the results we see for *Xlr4c* 4C domains do not suggest that X^P is not interacting with genes related to intelligence and cognition, rather that X^M appears to interact *more* than X^P.

Interestingly, among the *cis*-interacting regions of *Xlr3b*, I saw a similar more-varied range of phenotypes in regions unique to X^M compared to regions unique to X^P and regions found in both (Figure 1.3.16). However, there was little overlap to the phenotypes enriched in *Xlr4c* domains and none were related to intelligence or cognitive function. This is a curious result given the similarities of their long-range interaction profiles across the X-chromosome (Figure 1.3.4C).

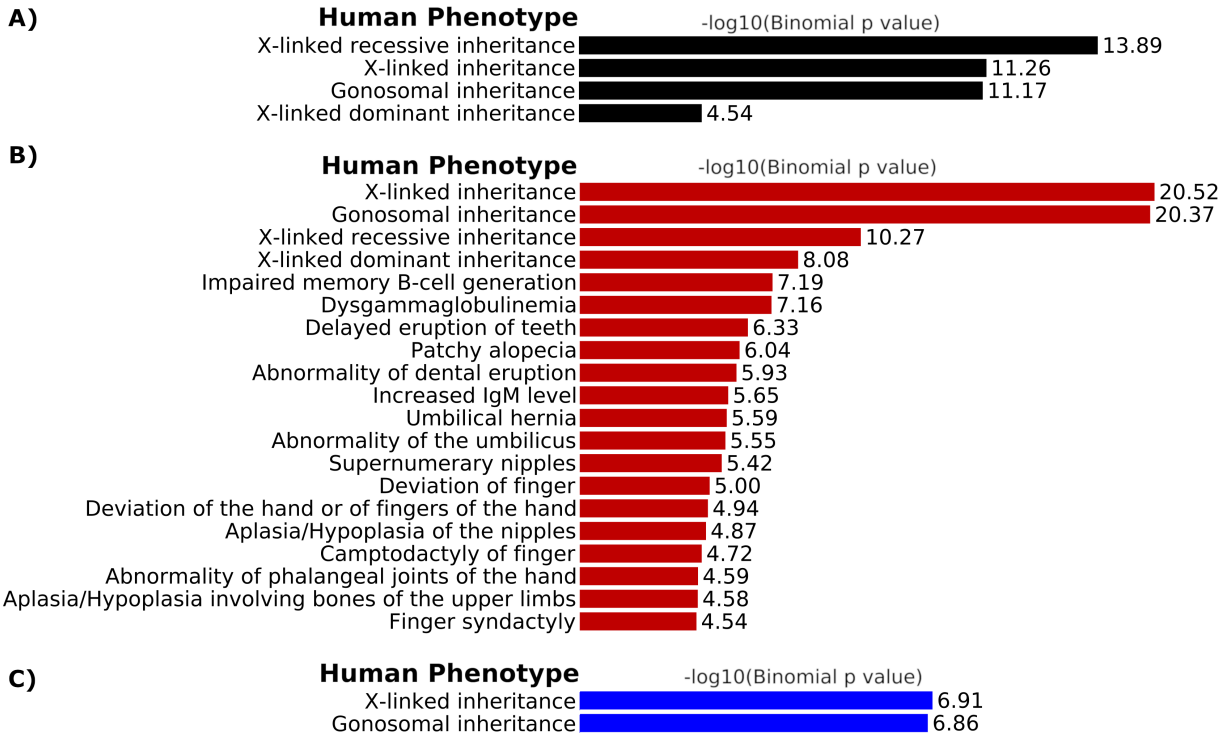


Figure 1.3.16: **Functional annotation of genes within *Xlr3b* 4C domains.** Prediction of functions of *cis*-regulatory regions common to X^M and X^P (A), found exclusively in X^M (B), and found exclusively in X^P (C).

1.4 Discussion

Autism spectrum disorder (ASD) is a complex developmental disorder that affects communication and behavior. As of 2018, one in fifty-nine children in the US are estimated to be diagnosed with ASD. Symptoms can include difficulty with communication and social interaction, restricted interests, and repetitive behaviors. The male bias of ASD describes a more prevalent diagnosis in males compared to females. This remains a poorly understood aspect of ASD, with statistics showing a male bias of ~4.3:1.

In this study, I have employed a "one-versus-all" technique to uncover all genome-wide interacting partners of *Xlr3b* and *Xlr4c*, two imprinted genes that have been implicated in ASD. Using X^M and X^P monosomic neonatal mouse brains for this analysis, I was able to compare the maternal and paternal interactomes genome-wide without any bias with regards to the singular X-chromosome. This study provides insight into key regions and genes that serve as the interacting partners to the imprinted *Xlr3/4* locus, contributes to a greater understanding of imprinted domains and their regulation, and raises important questions about the overabundance of genes related to cognitive function on the X-chromosome.

From the results, I saw that the distribution of 4C reads in *cis* is greater in X^M than X^P. In addition, the distribution of 4C domains is heavily skewed towards X^M, particularly on *trans* chromosomes. This could be because the *Xlr3/4* locus is active on the maternal allele, and thus, interacts with a greater number of genomic

partners. Such clusters of active genes and transcription factories may exist to provide coordinated expression of co-regulated genes.

Active regulatory regions are often characterized by open chromatin, nucleosome depletion, and DNase I hypersensitivity. In the case of X^M and X^P, 4C domains reside in open chromosomal compartments. This is evident in the earlier DNA replication timing of these regions. Additionally, 4C domains are enriched in RNA pol II occupancy, as well as active histone modifications such as H3K4me3, H3K36me3. The density of repetitive elements such as SINEs and LINEs also suggests that these 4C domains reside in gene-dense regions of the genome, in agreement with previously published 4C studies.

For many loci, the majority of meaningful interactions occur on the *cis* chromosome. Indeed, the most significant differences between X^M and X^P occur in a region directly surrounding the imprinted locus. These drastic differences likely represent differences in the chromatin landscape due to an imprinting mechanism, still not yet understood. As the region with the most significant differences is merely 2.7 Mb, genome-wide ChIP-Seq to characterize this region not ideal. Instead, a high-throughput method that allows for region selection, would be a good way to characterize the landscape surrounding this locus.

As the X-chromosome is highly enriched in genes related to intelligence and cognitive function, upon looking at the functional annotation of *Xlr4c* 4C domains, it came as no surprise that 4C interacting regions are enriched in these genes. Though both interact with genes implicated in intellectual disability, strikingly, it appears

that the maternal locus interacts with *more* of these partners than the paternal locus. Genomic elements that are unique to the maternal interactome include genes implicated in intellectual disability, behavioral/psychiatric abnormality, cognitive impairment, and autism spectrum disorder. This is a compelling result that may point to an overarching theme of the X-chromosome.

Is this a phenomenon that is unique to the imprinted *Xlr3/4* locus, which is known to contain differentially expressed genes? And what would it mean if it weren't? To validate these findings, 3D-FISH experiments could be employed. Another option is Hi-C analysis, which would require very high resolution, to capture differences between X^M and X^P. My 4C data suggests that the maternal *Xlr3/4* locus is in more frequent proximity to a certain subset of genes. These findings might contribute to the understanding of ASD and the male prevalence associated with it.

In my analysis, I have obtained a list of potentially significant regions and genes, many of which are implicated in intellectual disability and ASD. This data reveals many novel and relevant candidate genes for further analysis to uncover the biological significance of these findings. Further validation is needed to determine if this is a phenomenon that is reproducible with other loci on the X-chromosome. The following chapter will cover a similar interactome analysis of genes that are biallelically expressed and not known to behave differently between X^M and X^P. This is the first step towards understanding the maternal X susceptibility to cognitive and behavioral abnormalities, and it adds a layer of complexity to the epigenetic and three-dimensional regulation of imprinted gene expression.

Chapter 2

Epigenomic Profiling of the Maternal and Paternal X-Chromosomes

2.1 Abstract

Using 4C-Seq, I characterized the interaction profiles of six different genes implicated in intellectual disability and/or ASD across the X-chromosome. I found significant differences upon analyzing the interactome data with Genomic Regions Enrichment of Annotations Tool (GREAT), with the maternal X making far more contacts with other genes related to intellectual disability and cognitive impairment, in comparison to the paternal X. Using ChIP-Seq, I identified and annotated H3K27me3 peaks and Atrx binding sites. A total of 1,188 Atrx binding sites were called on the maternal X-chromosome, and 1,437 on the paternal. While a total of 2,499 H3K27me3 peaks were called on the maternal X-chromosome, and only 595 on the paternal. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to determine enrichment of biological terms. Altogether, significant differences were found between X^M and X^P datasets. The differences found here might indicate an important role in the maternal X susceptibility to cognitive and social impairment.

2.2 Background

Human sex determination, as with most mammals, follows the XX female/XY male chromosome system, where females are the homogametic sex, and males are the heterogametic sex [74]. Unique in the genome, these chromosomes form a non-homologous pair in half the population, representing an intriguing example of monoallelic gene expression. While dosage compensation mechanisms exist to prevent lethality, the hemizygosity found in males has led to extreme genetic consequences in terms of content, activity, and function of the genes on the X-chromosome [75]. The result is a large, gene-rich X-chromosome, and a small, gene-poor Y-chromosome that has and continues to degrade over time. With its atypical structure and biased gene content, the mammalian X-chromosome is truly distinct in its unique evolutionary trajectory.

In 1914, Muller proposed that the mammalian sex chromosomes differentiated from an ancestral autosome pair [76]. One of the pair acquired a sex-determining gene, *Sex Determining Region Y (SRY)*, and became the sex-specific partner. The accumulation of male-advantage genes around *SRY* led to repressed recombination between the two sex chromosomes, ultimately resulting in degeneration of what has become the Y-chromosome [75]. The human Y-chromosome has fewer than 50 remaining coding genes and shares a small pseudoautosomal region with the X [77].

In contrast, the X-chromosome shows an overall reduced rate of mutation, both in comparison to other chromosomes and across multiple species [78]. Comparative

gene mapping and cross-species chromosome paint studies have shown that the X-chromosome is almost entirely conserved in gene content. Species sequence variation is lower for X-linked genes than autosomal genes, and remarkably, evidence shows that the gene content of the X-chromosome bears a tendency towards specific functions, something that is not observed in the autosomes.

The X-chromosome is highly enriched in genes involved in sex and reproduction, termed sex- and reproduction-related (SRR) genes [79]. There is a high concentration of X-linked genes whose mutant phenotypes have effects on fertility or reproduction. In addition, genes related to intelligence and cognition are found in 5-fold excess on the X-chromosome. This high concentration of genes required for neural development and cognitive function results in a large number of X-linked mental retardation (XLMR) conditions found to localize to the X-chromosome. In addition to its unique gene content, the gene density of the X also differs from comparably sized autosomes. The X-chromosome possesses a gene density of about half that of an autosome, due to the much greater size of X-linked genes [80]. This oddity suggests that a large proportion of the X-chromosome is transcribed and generates numerous coding and non-coding regulatory RNAs, necessary for the complexity of the eukaryotic cell.

A model for the evolution of sexually antagonistic alleles was proposed to explain these unique characteristics of the X and Y-chromosomes [81]. Sexually antagonistic alleles describe alleles that are beneficial to one sex but are neutral or possibly detrimental to the other. As the X-chromosome is present in a single copy in males,

the hemizygous state will push for an accumulation of mutations that are advantageous to males, while deleterious effects in females would be masked owing to heterozygosity and mosaic expression. Thus, the accumulation of genes related to intelligence and cognitive development on the X-chromosome would suggest that intelligence is indeed a male-advantageous trait. Could alleles that enhance cognitive function allow males to mate more females and provide better for offspring? Interestingly, it seems that two different selection pressures result in the unique gene content of the X-chromosome.

Colloquialized as the ”brain and balls” theory, the intriguing coincidence of XLMR and SRR genes on the X-chromosome is backed by clinical evidence that mutations in these genes often affect cognitive performance as well as reproductive fertility. This suggests that these same genes have function in both brain and reproductive tissues, possibly because over the course of X-chromosome evolution, X-linked genes responded to two different selection pressures [82]. Natural selection for male reproductive fitness would select for SRR genes, while sexual selection would select for intelligence. Evidence suggests that as a result of this sex-specific selection, genes on the X-chromosome acquired brain-specific function during the evolution of placental mammals [83].

Cognitive impairment seen in patients with Turner syndrome and the male prevalence of ASD and XLMR conditions certainly suggest that the X-chromosome contains a disproportionate number of genes involved in brain development and function. The parent-specific bias of females with Turner syndrome leads to the question

of whether there are fundamental differences between the maternal and paternal X-chromosomes, such that individuals who solely inherit a maternal X-chromosome are more susceptible to cognitive and social impairment [8].

Nuclear architecture describes the complex relationship between the transcriptional activity of a gene and the spatial environment around it. Nuclear organization and the spatial segregation of active and inactive chromatin inside the nucleus can be necessary for proper cell function. The topology of the X-chromosome and its position inside the nucleus reflects its expression status in the context of dosage compensation and XCI [84].

To explore fundamental differences in chromatin architecture between the maternal and paternal X-chromosomes, I performed 4C-Seq using various viewpoints across the mouse X-chromosome to gain detailed insight into X-chromosome nuclear organization. This technique was previously used to demonstrate that the imprinted *Xlr3/4* locus, maternally expressed and paternally silent, displays different interaction profiles between maternal and paternal X-chromosomes. Now, using viewpoints across the X-chromosome that are not differentially expressed, I demonstrate that the maternal and paternal X-chromosomes nevertheless fold differently, with X^M being unique in a consistently greater number of interactions with a defined subset of other chromosomal loci–specifically, those related to intellectual disability, cognitive function and impairment.

In addition, I performed ChIP-Seq to analyze a histone modification, H3K27me3, and transcription regulator Atrx binding on the maternal and paternal X-chromosomes.

H3K27me3, a silencing modification associated with heterochromatin, serves as an imprinting mark for DNA methylation-independent autosomal imprinting and *Xist* for paternally imprinted XCI [6][7]. Atrx has also been implicated in the regulation of imprinted genes [67][85]. Functions of the Atrx chromatin regulator include a role in the postnatal silencing of imprinted genes, as well as facilitation of transcription elongation through G-rich regions of the genome.

H3K27me3 and Atrx were chosen for ChIP-Seq analysis because of recent studies that emphasized their involvement in the regulation of imprinted domains. The aim of this study is to search for differences in enrichment and binding between X^M and X^P. H3K27me3 domains and Atrx binding sites were annotated and analyzed for functional enrichment with the aim of characterizing differences across the X-chromosome. Taken together, this data unveils differences between the maternal and paternal X-chromosomes that are otherwise completely identical at the nucleotide level and suggests an explanation for the maternal X-related susceptibility to cognitive and social impairment.

2.3 Results

2.3.1 Interaction profiles of the maternal and paternal X-chromosomes

To characterize how interactions might be different between maternal and paternal X-chromosomes, I again utilized the 4C-Seq strategy on mouse neonatal cortex tissue to determine whether these chromosomes might fold differently within the nucleus. Using six different bait loci across the X, all of which are genes implicated in

intellectual disability and/or ASD, I sought to understand the network of interactions that occur on the X-chromosome. I hypothesized that the maternal X-chromosome might interact more frequently with genes linked to cognitive impairment, in comparison to the paternal X-chromosome.

The genes chosen as bait for this experiment are: *Syn1*, *Cul4b*, *Fmr1*, *MeCP2*, *Nlgn3*, and *Ap1s2*. Significant domains were determined using previously described methods. Figure 2.3.1 shows the significant interacting domains across the X-chromosome for all six viewpoints. For each viewpoint, the distribution of interacting domains spreads far across the chromosome in X^M and X^P, with most interactions occurring in gene-dense regions of the chromosome, consistent with existing 4C data. As the majority of interactions are well known to occur on the *cis* chromosome, and the X is unique in its gene content and function, I set out to characterize the functional annotation of the genes within these 4C domains.

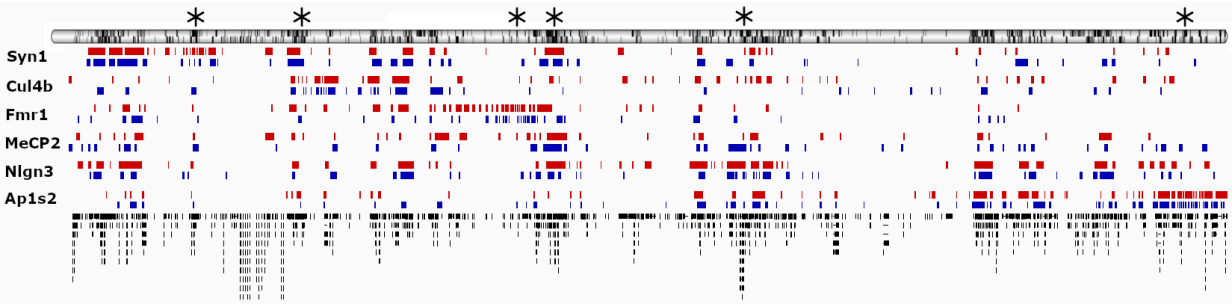


Figure 2.3.1: **Significant 4C domains of six viewpoints across the X-chromosome.** Distribution of significant domains in X^M (red) and X^P (blue). Bars representing long-range *cis* interactions from *Syn1*, *Cul4b*, *Fmr1*, *MeCP2*, *Nlgn3*, *Ap1s2* are shown above gene density (black). Asterisks indicate the position of the interrogated genes on the X-chromosome.

2.3.2 X^M 4C domains are enriched with genes involved in cognitive function

4C domains for each viewpoint were input to Genomic Regions Enrichment of Annotations Tool (GREAT) to analyze the functional significance of the interacting sequences. I observed significant enrichment of terms related to cognitive function in the human phenotype ontology and disease ontology (Table 2.3.2).

In X^M, *Syn1*, *Fmr1*, *MeCP2*, and *Nlgn3* all interact with genes related to intellectual disability and cognitive impairment, while in X^P, these genes do not show the same enrichment of terms. In *Cul4b*, both X^M and X^P were enriched for contacts with intellectual disability genes, while in *Ap1s2*, both X^M and X^P were enriched for contacts with intellectual disability, cognitive impairment, and behavioral/psychiatric abnormality. Interestingly, in X^M, *Fmr1* (p=3.8734e-5) and *Nlgn3* (p=5.7585e-7) interacting domains were enriched for genes that escape X inactivation, but not in X^P.

The disparity between X^M and X^P 4C domains is surprising given the similarity between their interaction profiles (Figure 2.3.1). One does not make more interactions than the other. Yet X^M seems to show more interactions with genes related to cognitive function, than X^P. For ease of subsequent analysis, all X^M 4C domains from the six libraries and all X^P 4C domains were pooled.

The pooled data shows great similarity between X^M and X^P. Figure 2.3.2 shows the number of associated genes per 4C domain. Most domains contain either one or two genes, with the majority containing the latter. Figure 2.3.3 shows the distribution of 4C domains binned by absolute distance to the nearest TSS. Again, X^M and X^P

share near identical profiles in their distribution of sites across the X-chromosome.

Table 2.3.2: Enrichment of cognitive function annotations

Library	**39,X^M**	**39,X^P**
Syn1	Behavioral/psychiatric abnormality (p=4.4517e-7)	N/A
	Intellectual disability (p=1.9906e-6)	
	Cognitive impairment (p=2.7288e-6)	
Cul4b	Intellectual disability (p=8.2905e-7)	Intellectual disability (p=9.6101e-7)
Fmr1	Intellectual disability, moderate (p=1.9117e-8)	N/A
	Developmental disease of mental health (p=3.15e-5)	
MeCP2	Intellectual disability (p=1.0848e-6)	N/A
Nlgn3	Intellectual disability (p=1.5112e-6)	N/A
	Cognitive impairment (p=4.3145e-5)	
	Abnormality of higher mental function (p=4.8503e-5)	
Ap1s2	Intellectual disability (p=5.9349e-7)	Intellectual disability (p=1.2615e-4)
	Cognitive impairment (p=1.7583e-6)	Cognitive impairment (p=7.0588e-6)
	Behavioral/psychiatric abnormality (p=5.8915e-6)	Behavioral/psychiatric abnormality (p=2.348e-6)

However, when it comes to the functional annotation of the domains, once again, I found that X^M has a far more interesting and wide-ranging profile than X^P. Figure 2.3.4A shows the enriched terms under the human phenotype ontology. X^M shows enrichment of behavioral/psychiatric abnormality (p=1.0671e-9),

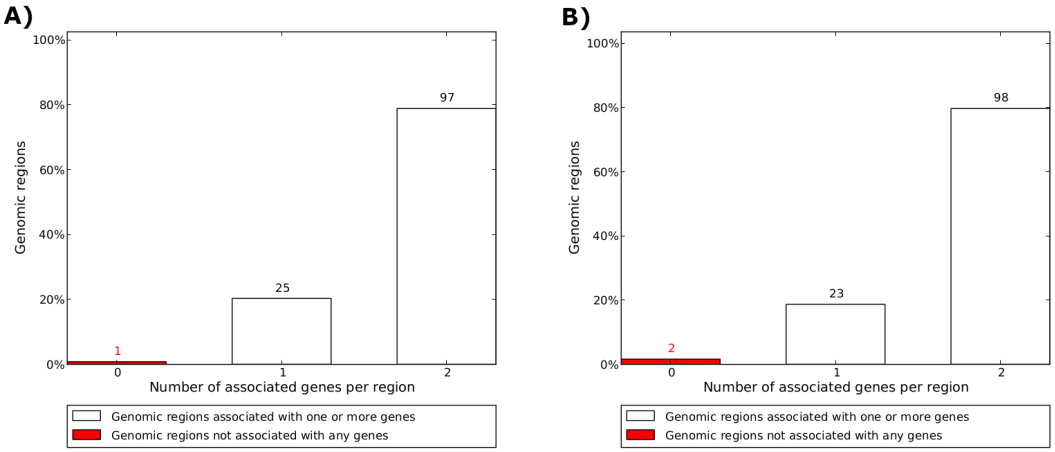


Figure 2.3.2: **Number of associated genes per 4C domain.** Number of associated genes per region in A) X^M and B) X^P.

cognitive impairment (p=1.6746e-7), and intellectual disability (2.8138e-7). In addition, autism (p=6.3355e-7) and autism spectrum disorder (p=7.7576e-7) were also enriched terms. X^P only showed enrichment of intellectual disability (p=3.9236e-6) and behavioral/psychiatric abnormality (p=6.5693e-6). Figure 2.3.4B shows the enriched terms under the disease ontology, where intellectual disability is an enriched term in both X^M (p=3.1407e-12) and X^P (p=8.5335e-8).

Figure 2.3.5 is a directed acyclic graph (DAG) representation of the human phenotype ontology enriched terms, where each node represents a single term, and the radius of the node correlates to the binomial fold enrichment. The hierarchical representation of terms highlights the differences between X^M and X^P. The node branch under abnormality of higher mental function in X^M contains cognitive impairment, intellectual disability, behavioral abnormalities, and autism spectrum disorder. In comparison, many of these terms are absent in X^P.

Next, to better understand why enriched terms might be so different between

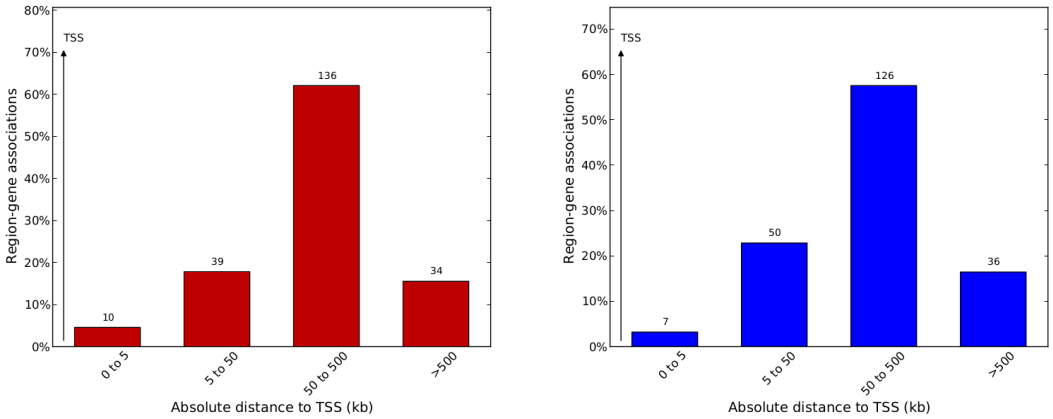


Figure 2.3.3: **Distribution of region-gene associations within 4C domains.** Distribution of 4C domains binned by absolute distance to TSS in X^M (red) and X^P (blue). Analysis done with Genomic Regions Enrichment of Annotations Tool (GREAT).

X^M and X^P, I looked to see if there was any enrichment in the MSigDB predicted promoter motif ontology and the MSigDB miRNA motif ontology. The predicted promoter motif ontology contains sets of genes that share a transcription factor binding site in their promoters, and the miRNA motif ontology contains genes that share a 3'-UTR miRNA binding motif. Using the same 4C domains as input, the only enriched motifs were found in the X^M domains (Figure 2.3.6). There were no enriched motifs found in X^P.

2.3.3 X-chromosome profiling of H3K27me3 and Atrx

To comprehensively profile parental origin-related differences of histone modification H3K27me3 and chromatin remodeler Atrx on the X-chromosome, I performed standard ChIP-Seq on 39,X^M and 39,X^P monosomic neonatal mouse brain. Replicate libraries were shown to be highly reproducible and data was pooled for subsequent analysis. Peaks and binding sites were screened using MACS2, a computational

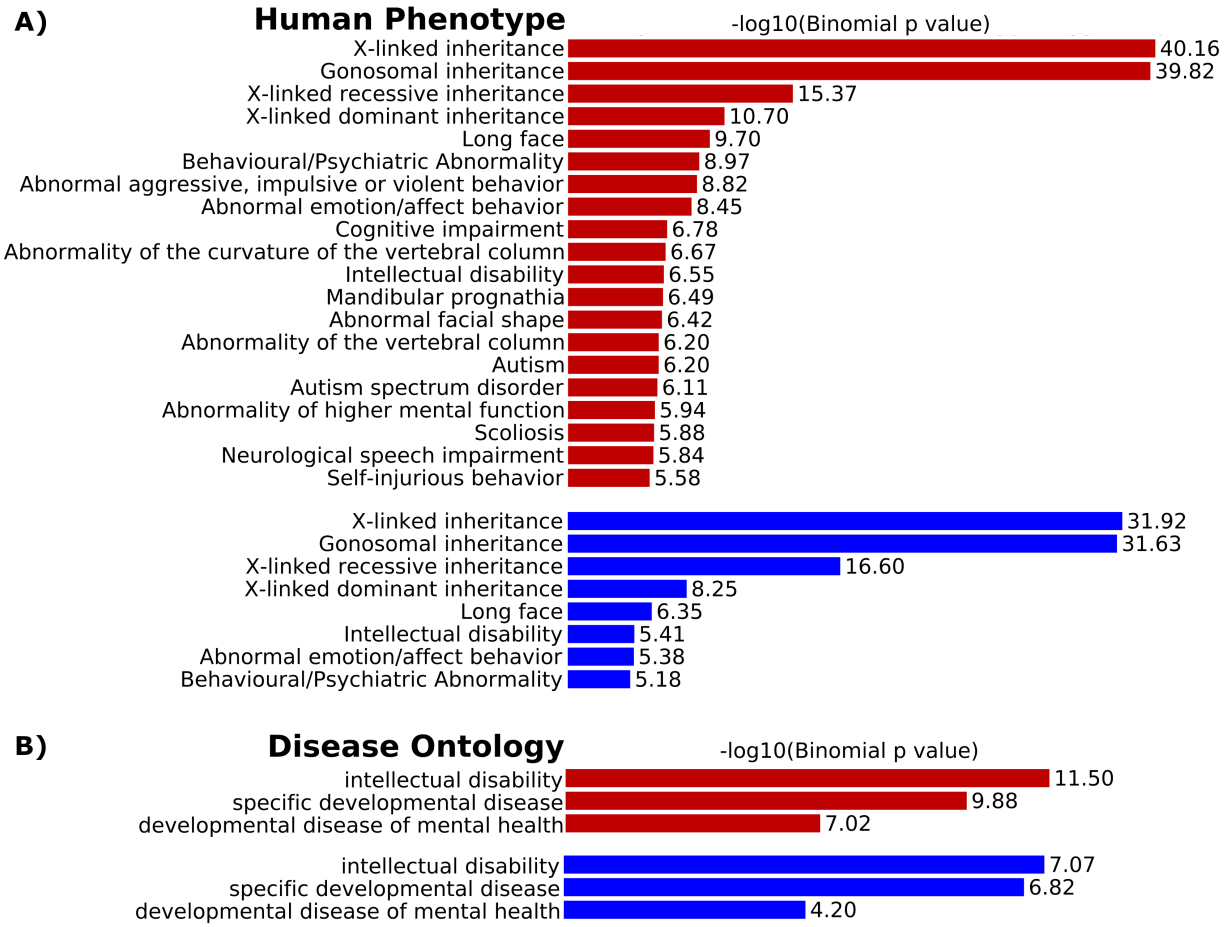
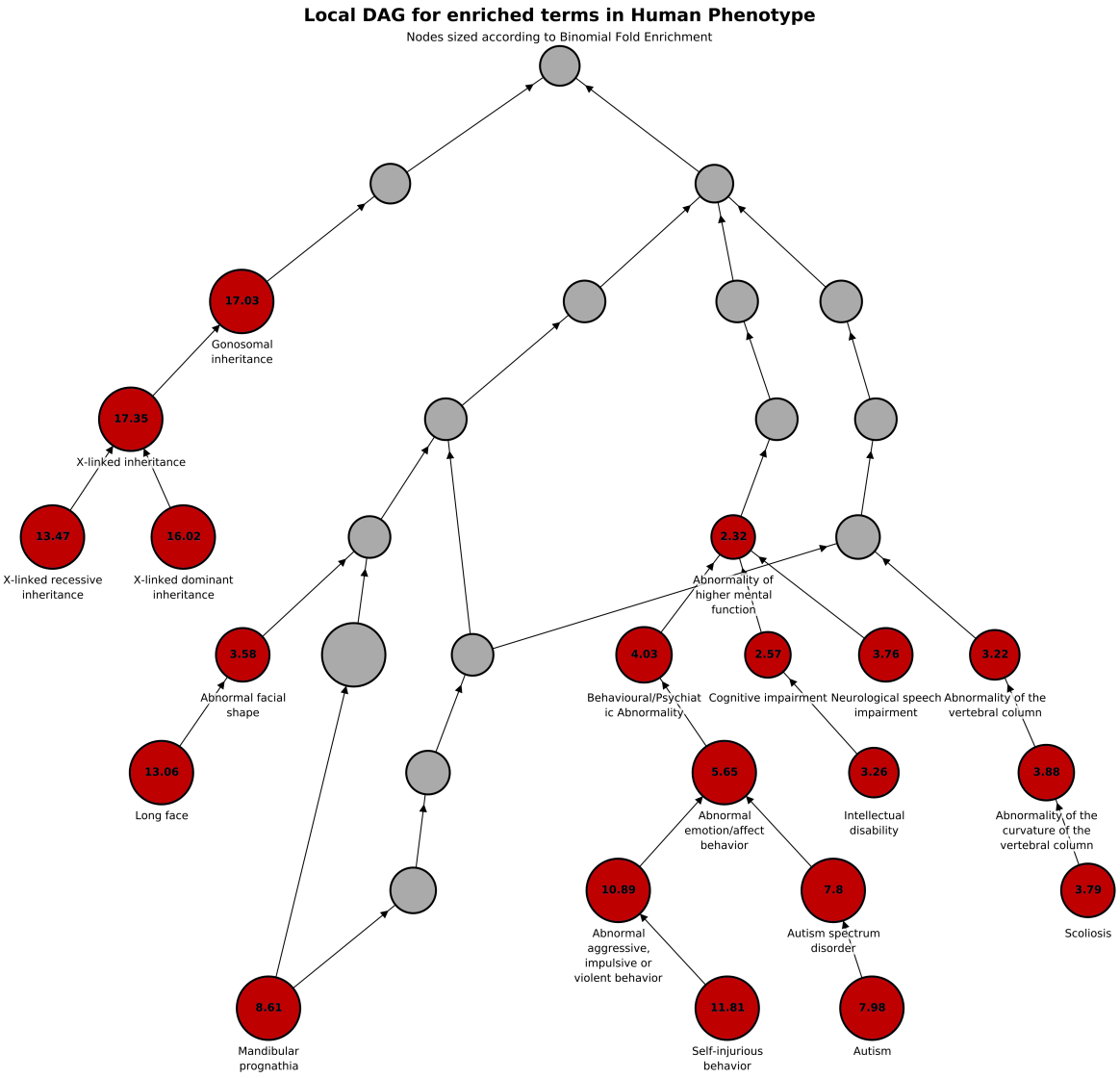


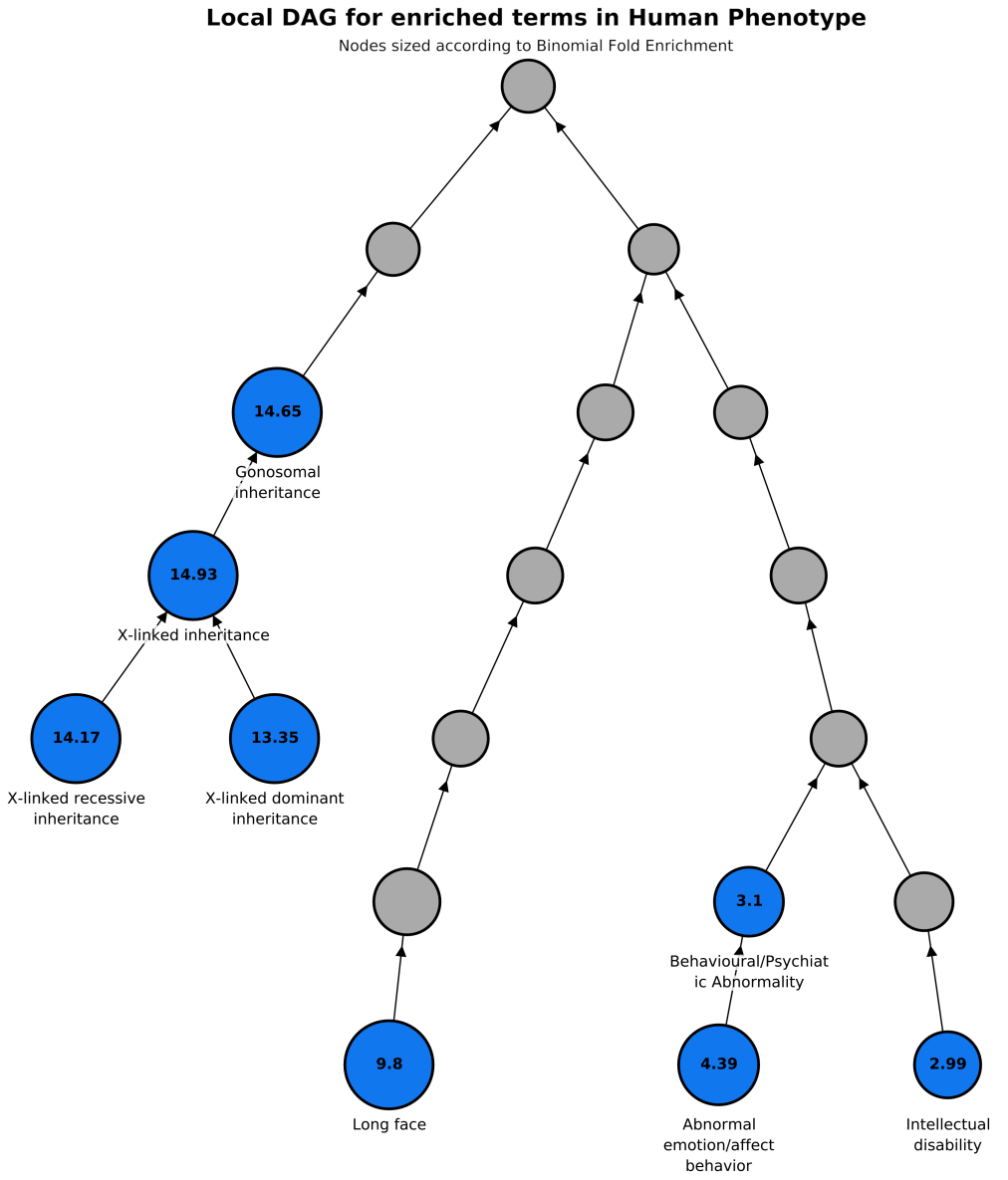
Figure 2.3.4: **Functional annotation of genes within 4C domains.** Prediction of functions of *cis*-regulatory regions in X^M (red) and X^P (blue). A) Human phenotype ontology. B) Disease ontology.

program used to identify genome-wide locations of protein-DNA interactions generated from ChIP-Seq data. Peaks were then annotated using ChIPseeker, an R package that supports annotation, visualization, and comparison between multiple ChIP-Seq data sets.



(a)

Figure 2.3.5: **Human phenotype functional annotation of 4C domains.** *Continued on next page.*



(b)

Figure 2.3.5: **Human phenotype functional annotation of 4C domains.** Human Phenotype ontology of X^M (red) and X^P (blue) visualized in hierarchy as a directed acyclic graph (DAG).

Figure 2.3.7 shows heatmaps of Atrx and H3K27me3 ChIP peaks surrounding existing genes in the mouse genome, specifically regions +/- 3000 bp to TSS. The differences across the two experiments are evident from the heatmaps of ChIP-Seq

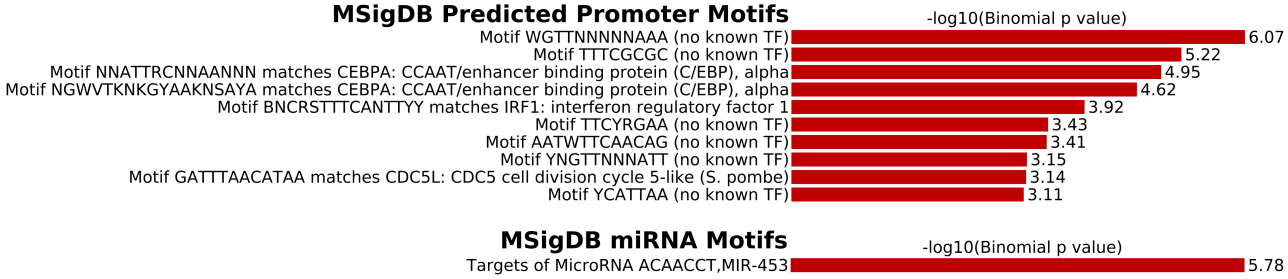


Figure 2.3.6: **Motifs within X^M 4C domains.** MSigDB predicted promoter motifs and miRNA motifs in X^M 4C domains.

tag enrichment. H3K27me3 shows a strong enrichment centered at the TSS, whereas Atrx shows no such localization to the TSS of genes. As H3K27me3 is a silencing histone modification, it follows that it would be found near the start site of certain genes. Whereas Atrx is a chromatin remodeling factor that binds throughout the genome, not necessarily at the beginning of gene bodies.

For comparison, Figure 2.3.8 shows heatmaps of ATAC-Seq tag density. ATAC-Seq is a molecular technique that shows open, accessible regions of the genome. As expected, its tag density shows a very strong enrichment at and near the TSS.

Next, I wanted to focus on differences between the maternal and paternal X-chromosomes. After extracting the peaks that are located on the X-chromosome, an averaged view of ChIP-Seq tag enrichment shows differences in peak count frequency between X^M and X^P (Figure 2.3.9). Whereas the bell-curve distribution of ATAC-Seq tag density shows relative similarity between the two, Atrx and H3K27me3 ChIP-Seq peaks show stark differences between X^M and X^P.

A total of 1188 Atrx binding sites were called on the maternal X-chromosome, and 1437 on the paternal. While a total of 2499 H3K27me3 peaks were called on

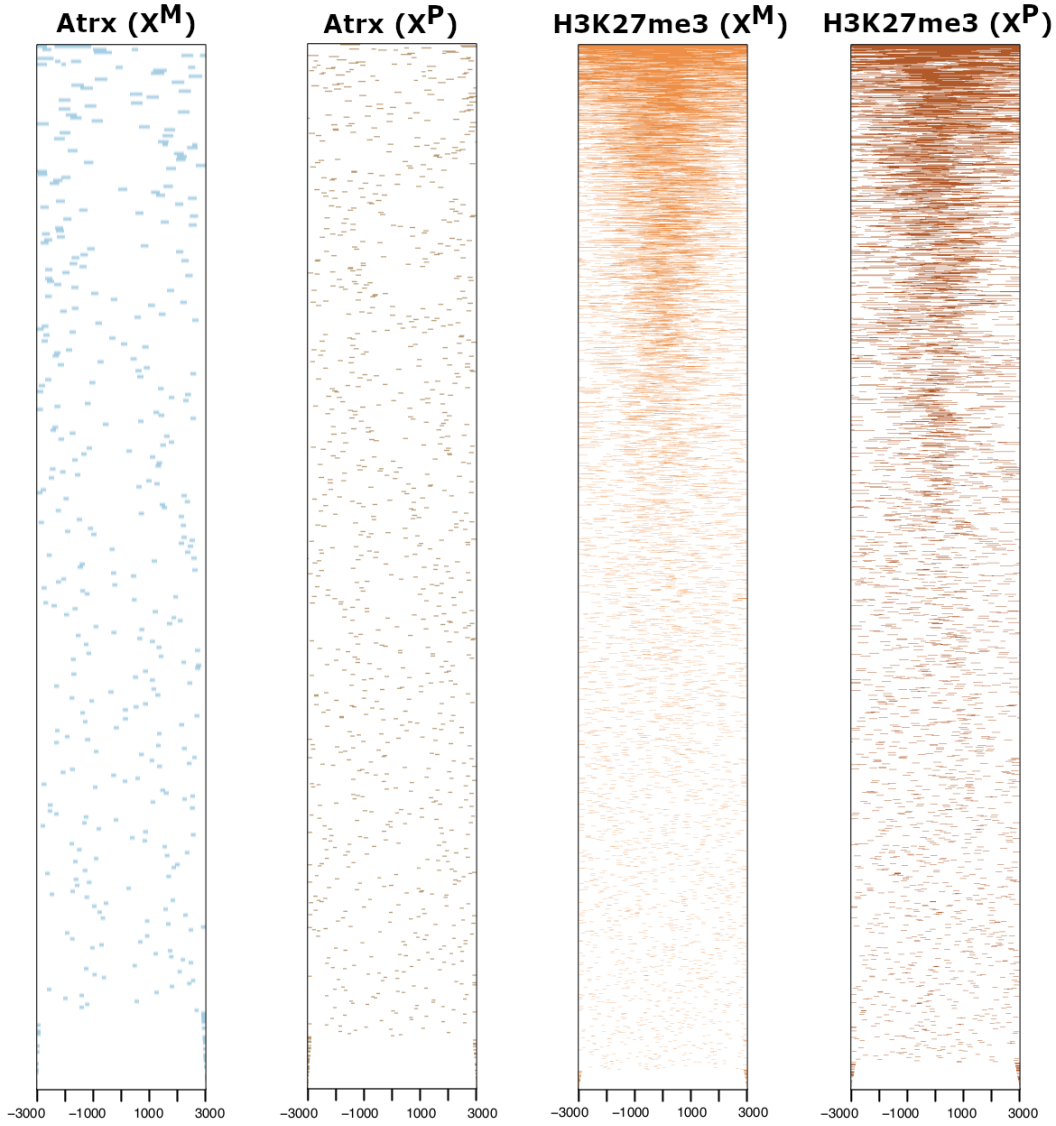


Figure 2.3.7: **Heatmap of Atrx and H3K27me3 ChIP-Seq peaks to TSS regions.** ChIP peak binding to TSS regions, +/- 3000 bp, visualized as a heatmap. Shows the distribution of reads in X^M and X^P.

the maternal X-chromosome, and only 595 on the paternal. The distribution and genomic annotation of these peaks across the X-chromosome can be seen in Figures 2.3.10 and 2.3.11.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was done on these peaks to search for enrichment of biological themes. KEGG is a collec-

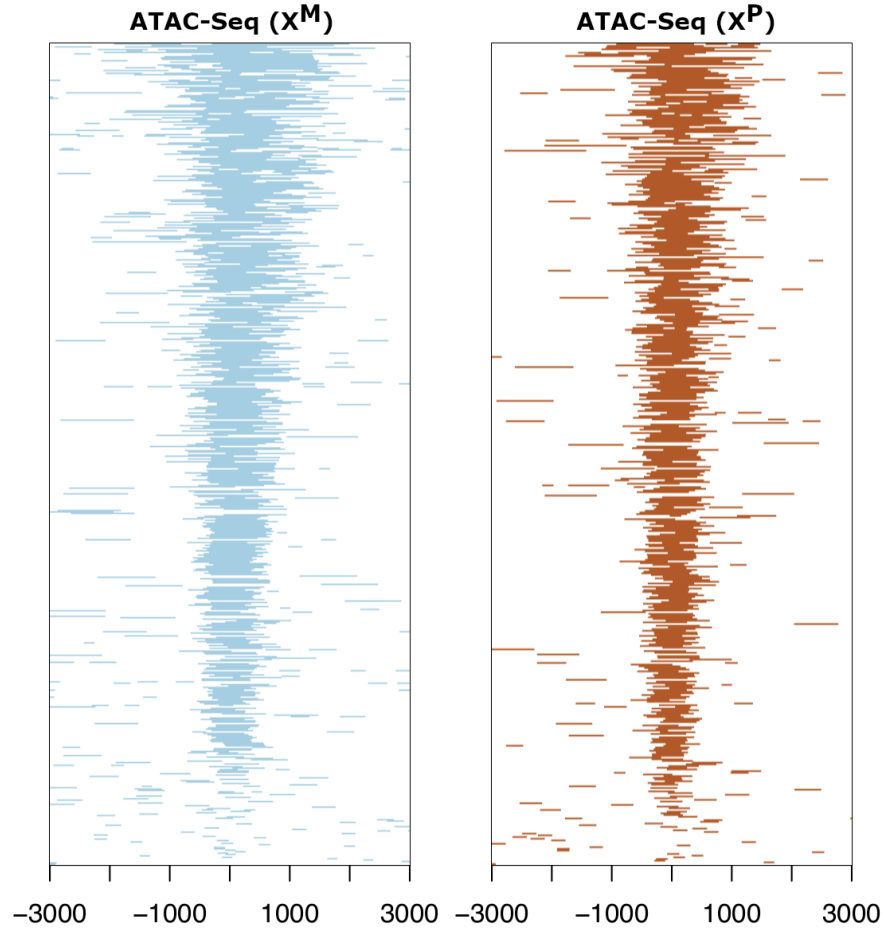


Figure 2.3.8: **Heatmap of ATAC-Seq peaks to TSS regions.** ATAC peaks surrounding TSS regions, +/- 3000 bp, visualized as a heatmap. Shows the distribution of reads in X^M and X^P. Data courtesy of Glenn Milton.

tion of databases that annotates genes to biological pathways. Within the Atrx datasets, KEGG pathways related to neuronal system and protein-protein interaction at synapses, among others, were found to be enriched (Figure 2.3.12). In addition, activation of NMDA receptors and postsynaptic events was solely enriched in X^M, and GABA A receptor activation was solely enriched in X^P.

Within the H3K27me3 datasets, KEGG pathway analysis did not reveal significant enrichment among peaks on the X-chromosome. A genome-wide analysis revealed

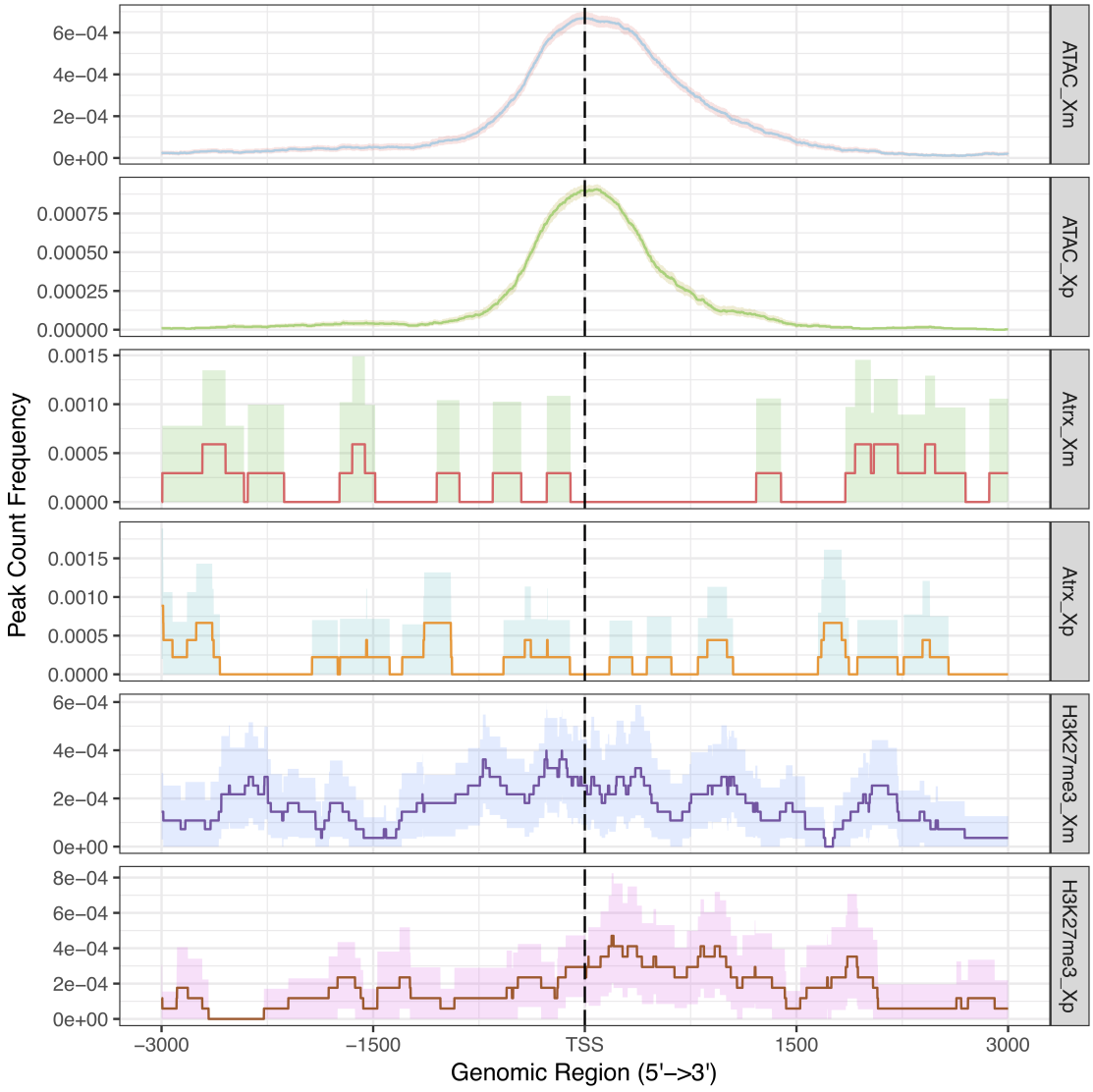


Figure 2.3.9: **Average profile of ATAC-Seq and ChIP-Seq peaks to TSS regions on the X-chromosome.** Average profiles of ATAC peaks and Atrx and H3K27me3 ChIP peaks to TSS regions on the X-chromosome. 95% confidence interval estimated by bootstrap method. Shows distribution of reads in X^M and X^P.

multiple pathways enriched among X^M and X^P (Figure 2.3.13). Many pathways enriched are signaling pathways, but some are related to cancer and transcriptional misregulation in cancer.

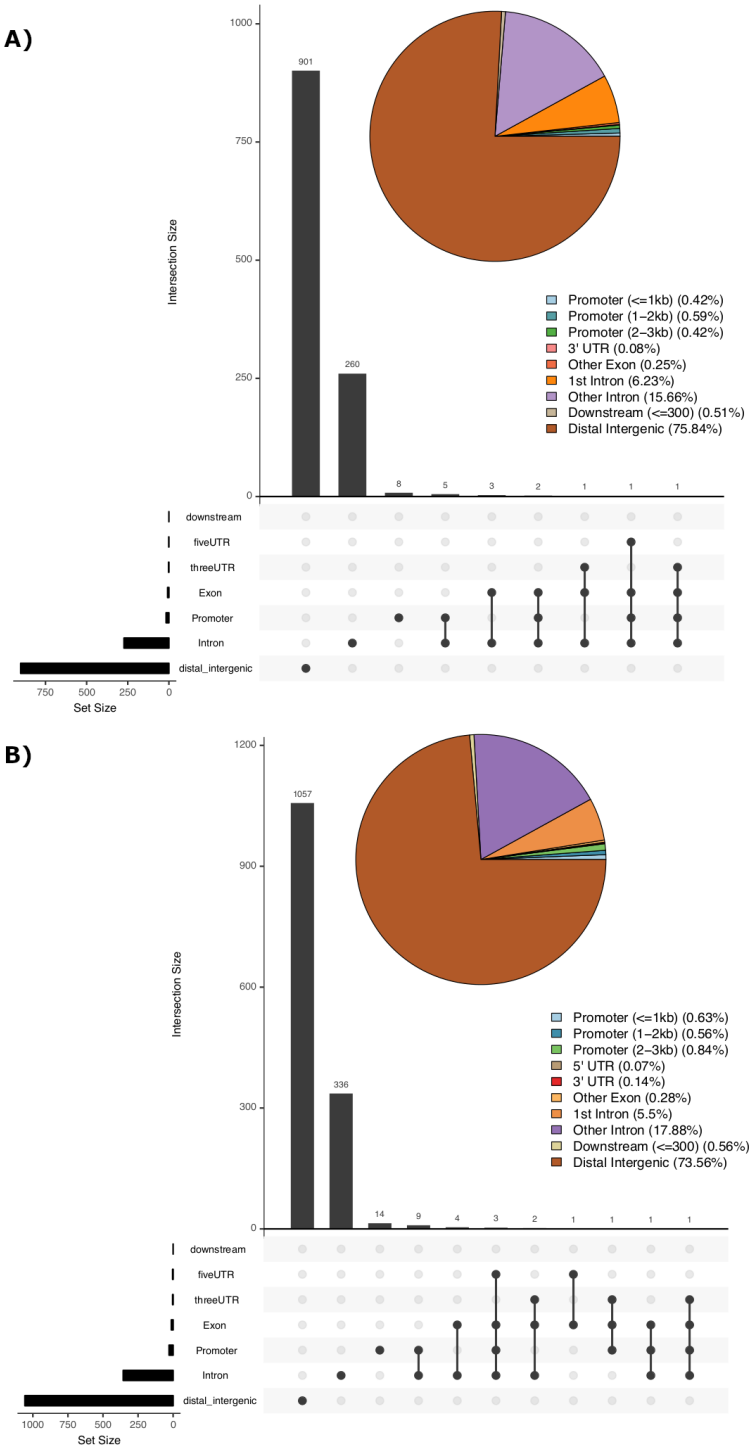


Figure 2.3.10: **Genomic annotation of Atrx binding sites on the X-chromosome.** Genomic annotation visualized as pie chart with full annotation overlap. Shows distribution of genomic elements in A) X^M and B) X^P.

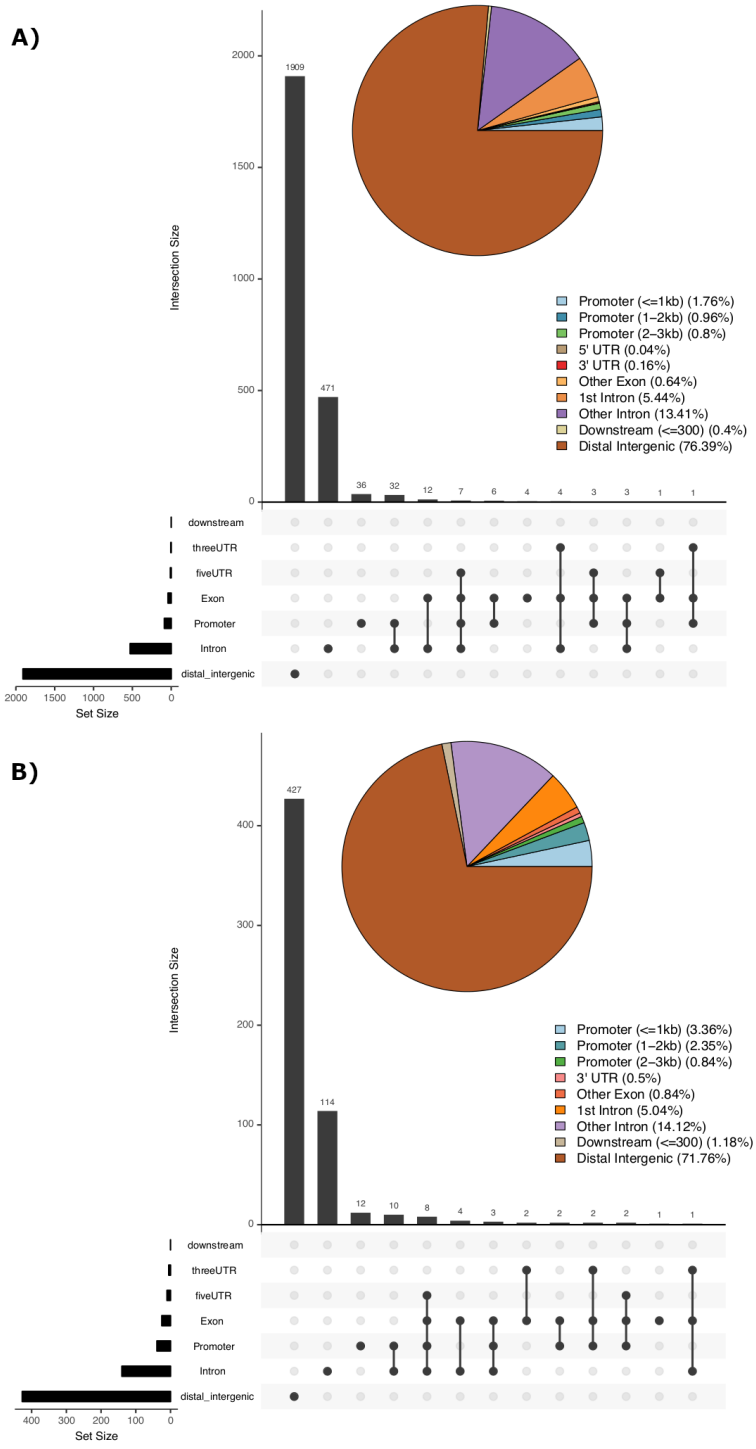


Figure 2.3.11: **Genomic annotation of H3K27me3 peaks on the X-chromosome.** Genomic annotation visualized as pie chart with full annotation overlap. Shows distribution of genomic elements in A) X^M and B) X^P.

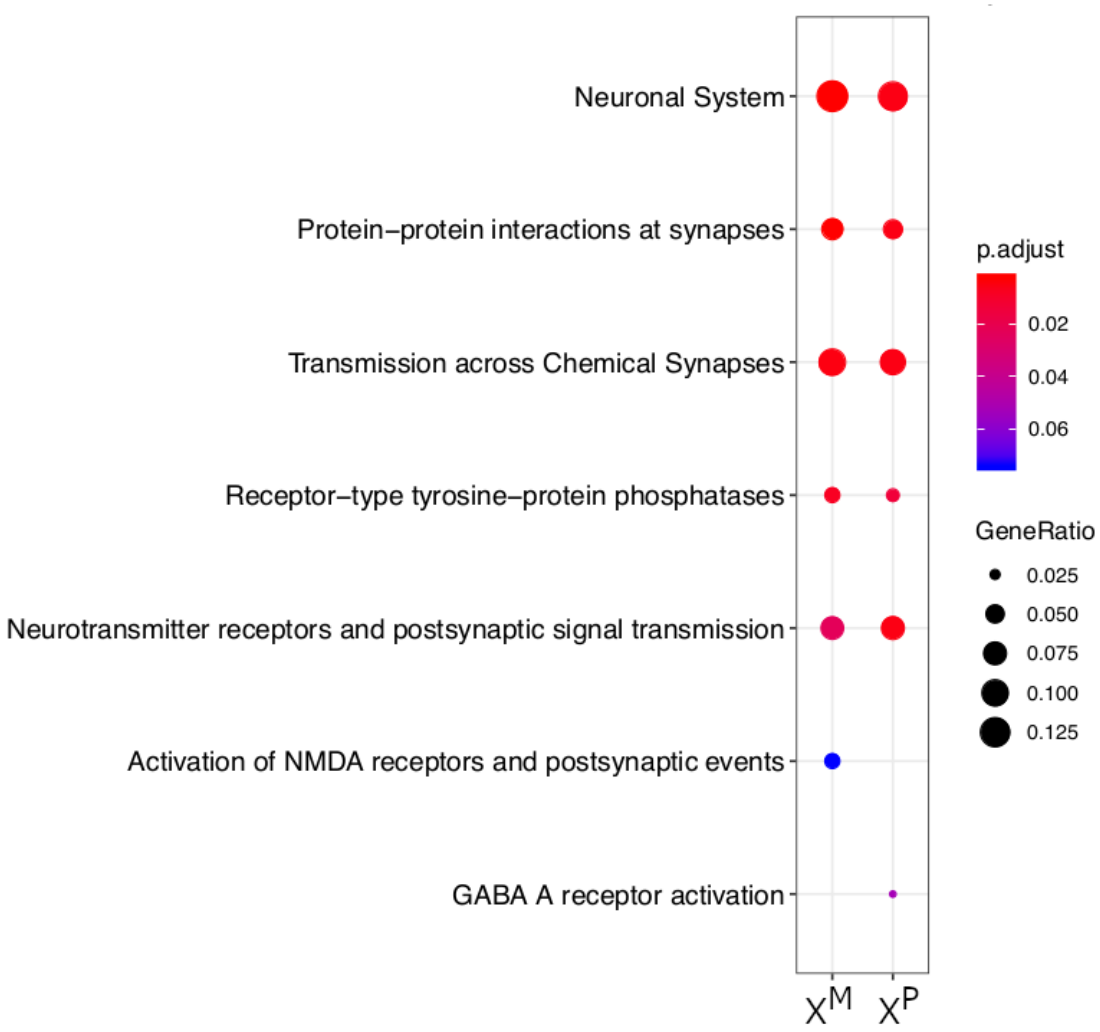


Figure 2.3.12: **Atrx KEGG enrichment analysis.** Dot plot of KEGG pathway enrichment analysis, comparing biological themes among gene clusters of X^M and X^P on the X-chromosome.

2.4 Discussion

In this study, I identified all interacting partners of six X-linked genes implicated in intellectual disability and/or ASD. I used 4C-Seq to identify interactions that occur in X^M and X^P cells and found differences in the functional enrichment of the genes within these interacting domains. Information extracted from these data sets suggests that the maternal X interactome contains a greater number of genes implicated

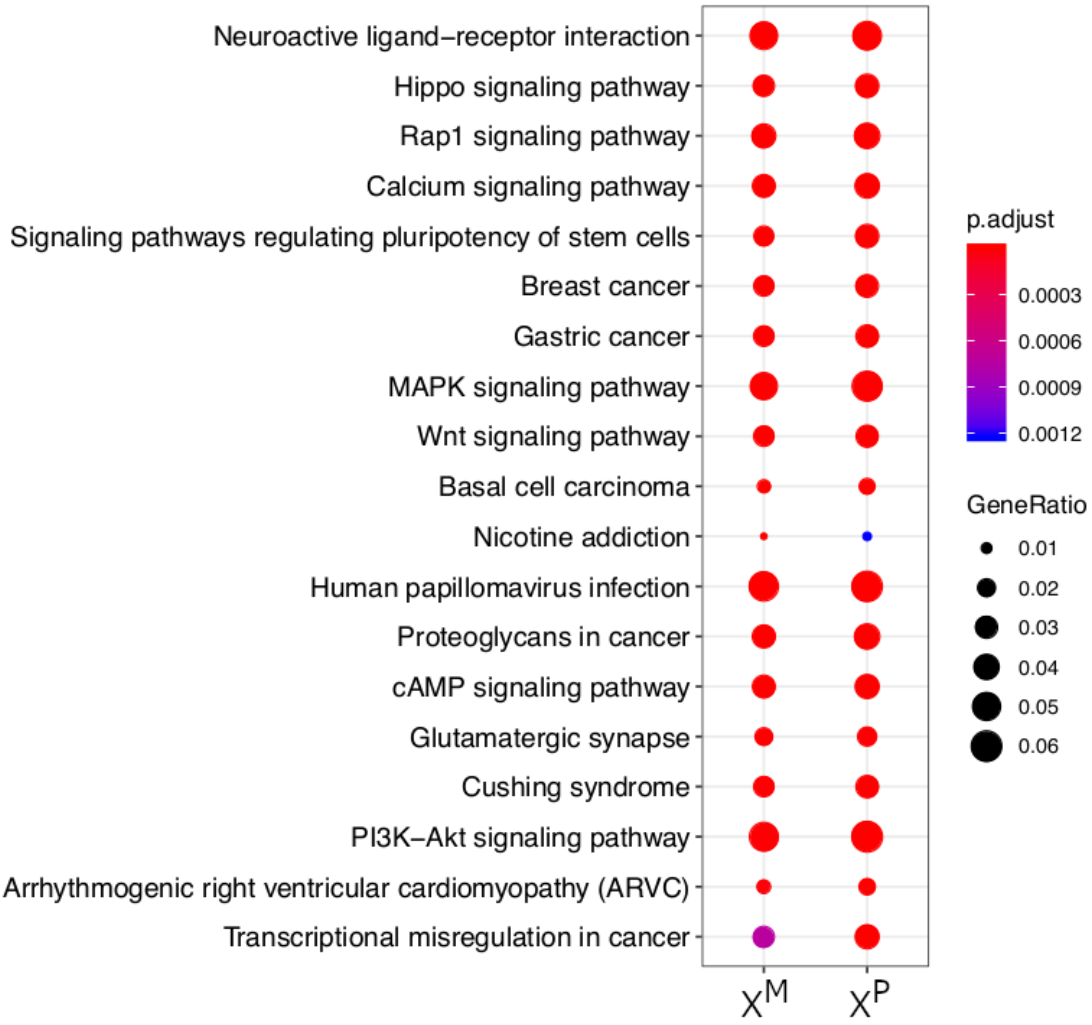


Figure 2.3.13: **H3K27me3 KEGG enrichment analysis.** Dot plot of KEGG pathway enrichment analysis, comparing biological themes among gene clusters of X^M and X^P, genome-wide.

in intellectual disability, cognitive impairment, and behavioral abnormality, in comparison to the paternal X interactome. Within the human phenotype ontology, the maternal X interactome is enriched with many more phenotypes, including autism and ASD. This conclusion is unexpected given the similarities in their distribution of sites, interaction profiles, and biallelic expression.

These results raise a number of important questions concerning the high propor-

tion of genes on the X-chromosome related to intelligence, cognitive function and development. Evidence for the maternal X susceptibility to intellectual disability, cognitive impairment, and behavioral abnormalities comes from the parent-specific bias of Turner syndrome and the male prevalence of ASD. While there is much overlap and similarity between the maternal and paternal 4C domains, the maternal interactome shows a much greater hierarchy of phenotypes that fall under abnormality of higher mental function. This data indicates that there may be a causal relationship between the chromatin topology of the X-chromosome and the maternal X susceptibility to intellectual disability. More evidence is needed to better understand the interplay between chromatin interactions and the cognitive phenotypes seen in Turner syndrome and ASD.

The potential implications of these results lead to a question of the biological significance of a wider and more varied maternal X interactome. Unlike the aforementioned *Xlr3/4* locus, the genes in this study are not imprinted, which makes the differences in long-range interactions all the more intriguing. Like other 4C datasets, I show that loci tend to interact with preferred neighbors in *cis*, usually among gene dense regions of the chromosome. Previous studies have also shown that on the X-chromosome, genes that escape X inactivation tend to contact each other in the nucleus [84]. This suggests evidence for genes interacting with other genes that share similar functions. Furthermore, these findings provide functional insight into genes on the X, implicated in intellectual disability and ASD.

A ChIP-Seq approach was used to analyze the histone modification, H3K27me3,

implicated in X-chromosome inactivation and methylation-independent genomic imprinting, as well as the transcriptional regulator, Atrx, also implicated in the regulation of imprinted genes. In this study, the ChIP-Seq peaks were analyzed with the primary focus of comparing the maternal and paternal X-chromosomes. While there were many overlaps in peaks between X^M and X^P, there were also peaks that were unique to one or the other. These results might be improved by a greater sequencing depth of the singular X-chromosome.

The average profiles of H3K27me3 and Atrx peaks across the X-chromosome show significant differences when it comes to peak count frequency across genomic regions near TSS. KEGG enrichment analysis also showed differences among the Atrx binding sites on the X-chromosome, but did not for H3K27me3. Among H3K27me3 domains, similar biological processes and pathways were found to be involved in X^M and X^P. This data suggests that there are some differences between maternal and paternal H3K27me3 modifications and Atrx chromatin remodeler binding sites. Further analysis is needed to determine the biological and functional significance of these differences.

Collectively, this data shows that chromosome contacts on the maternal X are biased towards genes related to intellectual disability and cognitive impairment. The X^M interactome making more promiscuous interactions is seemingly not conducive to neurodevelopment. Further experiments are needed to address the relationship between chromatin topology and the cognitive phenotypes seen in patients with ASD. Importantly, this data suggests that there are fundamental differences between

the maternal and paternal X-chromosomes, which are otherwise completely identical at the nucleotide level. This data provides evidence for parent-of-origin effects on chromatin interactions among genes and regulatory sequences, as well as the histone modification, H3K27me3 and chromatin remodeler protein, Atrx.

Chapter 3

Regulation of *Xlr3b* Gene Expression

3.1 Background

Imprinted genes are a subset of genes in mammalian genomes that are differentially expressed in a parent-of-origin dependent manner. Most autosomal imprinted genes exist in clusters and are collectively regulated by a single imprinting control region (ICR). The presence of local differential methylation is nearly ubiquitous in all imprinted clusters and is found to be necessary for maintaining parent-specific gene expression (PSGE). Reciprocally imprinted genes *IGF2* and *H19* are governed by a differentially methylated region (DMR) ~2kb upstream of *H19* that exhibits paternal-specific methylation. A deletion of this region abolishes PSGE, making this region an ICR [86]. Soon after, it was shown that zinc-finger binding protein CTCF binds to the *H19/IGF2* ICR in a methylation-dependent manner [87]. CTCF acts as an insulator protein, allowing downstream enhancers to activate the maternal *H19* allele. Lack of CTCF on the paternal allele induces a long-range chromatin interaction, bringing enhancers into proximity with the *IGF2* promoter.

In 2010, Kernohan *et al.* demonstrated that, in addition to CTCF, chromatin regulators ATRX, MeCP2, and cohesin also co-localize at the *H19* ICR, with preferential

binding on the maternal allele [67]. ATRX belongs to the SWI/SNF family of chromatin remodeling proteins and is implicated in ATR-X syndrome. Loss of ATRX in mouse forebrain tissues saw altered enrichment of CTCF and cohesin at the *H19* ICR and increased *H19* expression, without affecting DNA methylation on the paternal allele. ATRX null samples also displayed increased H3 and H4 acetylation, suggesting that loss of ATRX induces a more open, accessible chromatin state. These same chromatin modifiers were also found to co-localize at the ICR of the *Gtl2/Dlk1* imprinted cluster [67]. Similarly, loss of ATRX reduced occupancy of CTCF and SMC1 at the *Gtl2/Dlk1* DMR. It is important to note that the increased *H19* expression was maternal in origin, thus displaying no loss of imprint. Twelve other imprinted genes were found to display increased expression in ATRX null forebrain, providing evidence that ATRX may play a role in postnatal silencing of imprinted genes.

ATRX partners with DAXX, a H3.3-specific chaperone, to facilitate the deposition of the H3.3 histone variant at telomeres and pericentric DNA. ATRX also has genome-wide targets, binding to euchromatic regions containing CpG islands and/or G-rich tandem repeat sequences predicted to form G-quadruplex (G4) structures, to regulate local gene expression [88]. G-quadruplex structures are associated with inhibition at gene promoters and promoter-proximal pausing. Levy *et al.* found that ATRX binds within regions of the gene body to facilitate deposition of H3.3 and transcription elongation through G-rich sequences [68]. In the absence of ATRX, they saw increased RNA pol II occupancy along the gene body, indicative of polymerase stalling at G-rich regions, implicating ATRX in a mechanistic model that promotes

gene expression. A few years later, Voon *et al.* demonstrated that ATRX and H3.3 preferentially localize to the methylated allele of imprinted DMRs, and loss of ATRX led to loss of both H3.3 and H3K9me3 at intragenic CpG island targets, ultimately resulting in the de-repression of multiple imprinted genes [85].

ATRX is far from the only transcriptional regulator often assigned contradictory roles. CTCF is a highly conserved transcription factor that can function as a transcriptional activator, repressor, or insulator protein. CTCF also serves as one of the core architectural proteins, along with the cohesin complex, that helps establish the three-dimensional organization of the genome. Mutation of CTCF binding sites at the *H19* ICR results in loss of CTCF binding and *de novo* methylation of an upstream CTCF target site within *Igf2* DMR1, showing that CTCF can coordinate regional epigenetic marks [38]. MeCP2, a protein implicated in Rett syndrome, belongs to a family of methyl-CpG-binding domain (MBD) proteins. Curiously, this protein with an affinity for methylated DNA, can bind methylated and non-methylated sequences [89]. MeCP2 is highly expressed in brain, was found to be able to mediate nucleosomal compaction, and demonstrated a global role in regulating neuronal chromatin structure [90][91]. A newer hypothesis suggests MeCP2 may play a role in forming DNA loops, selectively looping methylated sequences in imprinted regions, thereby indirectly regulating PSGE [48]. Kernohan *et al.* showed that MeCP2 is required for ATRX recruitment and chromatin looping at the *H19/Igf2* locus. Deficiency of either ATRX or MeCP2 causes decreased frequency of long-range chromatin interactions, relating to increased nucleosome occupancy and decreased CTCF occupancy at

CTCF binding sites [92]. Altogether, their data suggests that MeCP2 recruits ATRX to control local nucleosome positioning, stabilize CTCF binding, and maintain higher order chromatin structure at multiple imprinted domains.

In addition to its role regulating imprinted domains, ATRX marks the inactive X-chromosome (Xi) in somatic cells and in cells showing imprinted XCI [93]. ATRX co-localizes with macroH2A, a marker of the Xi, and is retained throughout cell division and differentiation, suggestive of a potential role for ATRX in the maintenance of XCI. Furthermore, ATRX functions as a RNA-binding protein that directs Polycomb Repressive Complex 2 (PRC2) to the long non-coding *Xist* RNA [94]. PRC2 is a key factor necessary for methylating H3K27me3 and establishing repressive chromatin on the Xi. H3K27me3 acts as the imprinting mark of *Xist* and more recently, has been shown to control DNA methylation-independent autosomal imprinting [6][7]. ATRX depletion results in a lack of Xi-wide spreading of PRC2 and an inability to initiate XCI.

Atrx knockout in Sertoli mouse cells saw a delay in the onset of spermatogenesis, along with a range of spermatogenesis defects in adult mice [95]. This study looked at the role of *Atrx* in testis development and function, and found *Atrx* to be highly expressed in Sertoli cells, where it activates expression of *Rhox5* and other androgen-dependent genes.

Recently, Shioda *et al.* reported that *Xlr3b* is one of the most upregulated genes following an *Atrx* knockout in mouse hippocampus [96]. The increase of *Xlr3b* expression upon *Atrx* knockout is specific to the brain, and was not found in other

tissues analyzed. In addition to increased expression and RNA pol II occupancy, the level of 5-mC at *Xlr3b* as determined by bisulfite sequencing analysis was significantly decreased in the *Atrx* knockout. They found that a G-rich region, capable of forming parallel G-quadruplex structures, upstream of the *Xlr3b* transcriptional start site (TSS) was enriched for ATRX, DNMT1, DNMT3A, DAXX, and H3.3 in wild type mouse hippocampus. A loss of such enrichments upon *Atrx* knockout suggests ATRX binds to G-quadruplex structures in *Xlr3b* with DNMTs, DAXX, and H3.3, where it regulates DNA methylation and *Xlr3b* gene expression.

It was previously reported that either DNMT1 or DNMT3A knockout also promotes increased *Xlr3b* expression [39][40]. These results and prior work done by members of M. O'Neill lab suggest that both alleles are likely subject to the silencing effects of DNA methylation. Bisulfite sequencing failed to find differential methylation at the G-rich region upstream of the *Xlr3b* TSS, the same region where Shioda *et al.* reported decreased methylation post *Atrx* knockout.

For a while, it was believed that RNA pol II binds equally to the maternal and paternal alleles of the *Xlr3b* TSS in neonatal brain, and that a stalling mechanism within the gene body is responsible for the imprinted expression of these genes. Here, using ChIP-qPCR, I present new evidence that is both conflicting and inconclusive, in an attempt to further characterize the epigenetic environment at the so-called "locus of madness".

3.2 Results

Evidence for co-transcriptional gene regulation began with analysis of the *Xlr3b* primary transcript. Seth Kasowitz reported equal expression levels at the 5′ end of the gene, and only maternal expression at the 7th intron towards the 3′ end of the gene, suggesting that some mechanism of paternal silencing occurs during transcription. Soon after, Sohaib Qureshi analyzed RNA pol II binding in ChIP-qPCR experiments. He saw equal enrichment of the serine 5 phosphorylated form of RNA pol II (Ser5), which is detected during early transcription initiation, in agreement with results showing equal transcript abundance at the 5′ end of the primary transcript.

Qureshi's analysis with the serine 2 phosphorylated form of RNA pol II (Ser2), which is detected during transcription elongation, showed paternal enrichment at intron 3/exon 4, and maternal enrichment at intron 7. These results coincided nicely with H3K36me3 ChIP-qPCR results–a hallmark of active transcription–showing similar enrichment between alleles, which further supported a stalling event on the paternal allele within the *Xlr3b* gene body. RNA pol II 4H8 ChIP-Seq, also done by Qureshi, showed similar results–but shifted down the gene. His results showed paternal enrichment over an ultra-conserved region in intron 6 (instead of intron 3/exon 4), and maternal enrichment near the 3′ end of the gene (instead of intron 7).

However, upon attempting to replicate his data, I found unequal enrichment of RNA pol II at the *Xlr3b* TSS (Figure 3.2.1). ChIP was performed using an antibody

against RNA pol II Ser5 on X^M and X^P mouse neonatal cortex. I saw maternal enrichment relative to paternal at both the *Xlr3b* promoter and TSS, suggesting that transcription initiation and early elongation do not occur equally on both alleles. The Ser5 profile is considerably higher around the TSS where transcription is initiated.

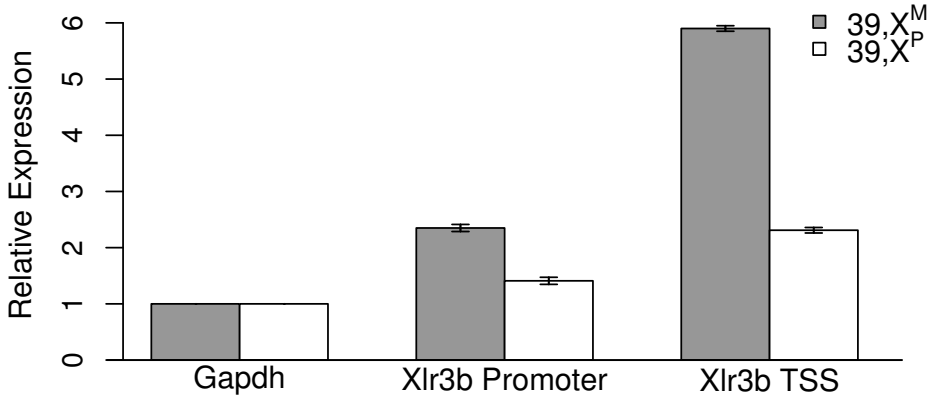


Figure 3.2.1: **RNA pol II enrichment at the 5′ end of *Xlr3b*.** RNA pol II Ser5 ChIP-qPCR on neonatal brain, showing enrichment of maternal RNA pol II at the 5′ end of *Xlr3b*, normalized to *Gapdh*. N=3.

Furthermore, upon tracing RNA pol II enrichment along the *Xlr3b* gene body using RNA pol II Ser2, I again found maternal enrichment relative to paternal throughout the gene body with the greatest difference at the 3′ end of the gene (Figure 3.2.2). This data is not indicative of a stall during elongation on the paternal allele. Instead, it suggests unequal initial binding of the polymerase to the TSS, culminating in a greater maternal buildup of polymerase towards the 3′ end of the gene.

Unfortunately, I was not able to replicate Qureshi's data regarding RNA pol II binding. This discrepancy could result from imprecise real-time measurements due to complications of PCR primer design within the region of interest. Primer pairs for RT-PCR must contain *Xlr3b*-specific SNPs to distinguish *Xlr3b* from its

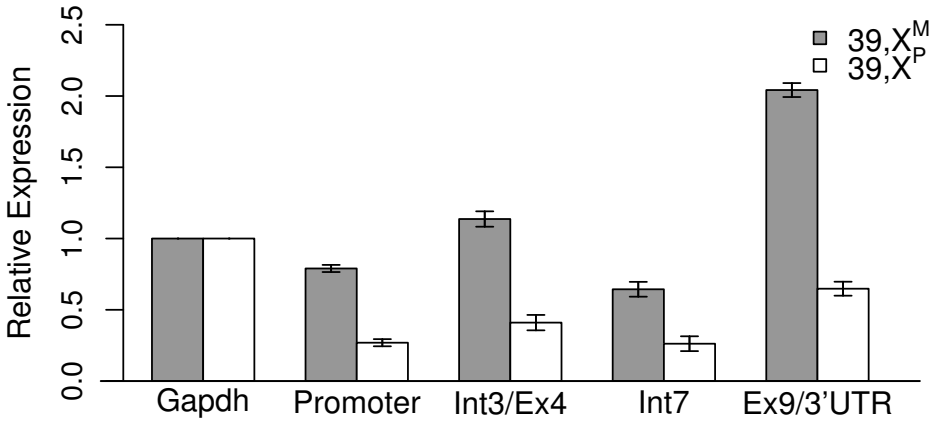


Figure 3.2.2: **RNA pol II enrichment across *Xlr3b*.** RNA pol II Ser2 ChIP-qPCR on neonatal brain, showing enrichment of maternal RNA pol II across *Xlr3b*, normalized to *Gapdh*. N=3.

non-imprinted paralogs, *Xlr3a* and *Xlr3c*. As a result, very few locations could be analyzed, and of the PCR products produced, some of Qureshi's primer sets (and my own) produced qPCR products past the boundaries of acceptable product size for real-time analysis, which could result in unreliable and highly variable data.

Next, I set out to confirm Shioda's result, which saw Atrx enrichment at a region in *Xlr3b* strongly predicted to form G-quadruplex structures (Figure 3.2.3). Shioda's analysis was done on hippocampus of P90 male mice. Their novel finding was that upon *Atrx* knockdown they saw an increase of *Xlr3b* expression and a loss of methylation. I performed ChIP-qPCR using the same Atrx antibody on neonatal cortex of X^M and X^P mice, to see whether a difference in expression is correlated with Atrx enrichment.

My results indeed show Atrx enrichment at the *Xlr3b* TSS–the very same region that is strongly predicted to form G-quadruplex structures. I saw greater Atrx enrichment at this region compared to further downstream at intron 7 (Figure 3.2.4),

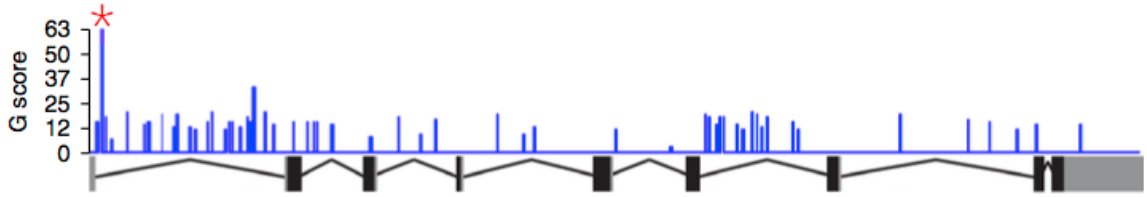


Figure 3.2.3: **Predicted G-quadruplex structures at *Xlr3b*.** Plot of G-scores (blue) indicating the likelihood of G-quadruplex formation along the entire *Xlr3b* genomic DNA sequence. Highest likelihood of G-quadruplex formation shown by red asterisk. Adapted from Shioda *et al.* [96] and QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php).

which falls in line with Shioda's results. However, I found enrichment of Atrx to be inconsistent across biological replicates. While there is slight enrichment of Atrx in X^P over X^M, it is not a statistically significant difference. As *Xlr3b* expression is reduced on the paternal allele, greater enrichment of paternal Atrx would be expected, according to Shioda's model. Curiously, *Atrx* knockdown increases *Xlr3b* expression–just like it does *H19*. Yet *Atrx* also promotes transcription elongation through G-rich sequences–which exist in *Xlr3b*. Given the inconsistency between biological samples and the contradictory roles of this chromatin remodeler, the question of whether *Atrx* plays a role in regulating *Xlr3b* imprinted expression is still yet to be answered.

If differential Atrx binding is occurring at the *Xlr3b* TSS, it is possible that the presence of the chromatin remodeler binding to the DNA could have affected the 4C technique using baits designed in this region. Figure 3.2.5 shows the location and sequence of the two pairs of 4C primers designed near the *Xlr3b* TSS. The HindIII primers encompass a 366 bp region from primary to secondary restriction site, and

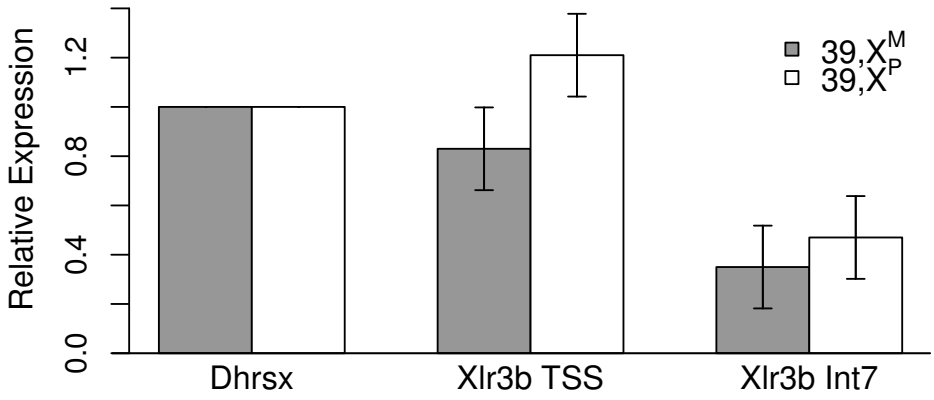


Figure 3.2.4: **Atrx enrichment at the *Xlr3b* TSS.** Atrx ChIP-qPCR on neonatal brain, showing slight paternal enrichment of Atrx at the *Xlr3b* TSS, normalized to *Dhrsx*, a positive control for Atrx binding. N=3.

are designed upstream of the region Shioda *et al.* suggests Atrx is binding. The DpnII primers encompass a 502 bp region and are designed directly on top of the region of Atrx binding. This could potentially explain the differences between *Xlr3b* and *Xlr4c* DpnII libraries, as the latter may not be affected by Atrx binding.

```
   1 ACCTTTGTGGAATTTCCTCTACAGCTTCACCTGTGTGCTGGTTTGTATGAGATGTGTCCCATTGACTCGGACATTTGAATACTGTATC
  89 CCCTGTTGTGTAGATGTTAGAATGTCCTGGCTGTAGAAAACCAGAAGTCTCCCAATACTGTCTTGTTCTCATCGCCCCAACTTGTAGG
 177 GGTCAGCCACATCCTCGGAAAGGATGCACCTCTTTGGAAAGGCTGGGACTAAGTAGCTGTGACTGCTCTCTCAAGGGAAGGATGCAGT
 265 CAGTGGCTCAGTGGTGTGGTGGGTTTGGTTTGCCGAGGTGAAGTACTGGGTGAAAGGGCATCCTCAGCTAAGCTGAGTTAAGAAGAGT
 353 AATGCTGGGACCACAGTCTTTGGGAGAGGATGGTGTGGATCACAAGGAGACTTGCCTAAGAGGGCCTGTGATGTAATGTGCCTGAGTC
 441 TGGCTTTAGCCTCTGACTCGACTAGTAACCCCTGGAGACTCATTGCAGAAGTTGCAGAGATAGAGAGCCTGGGATGTCTGAGCCGGAC
 529 ACTGCGGTCAAGGTACGGGCTCATAGTATCCAACATCTGCCCGTTTTGTACTGTTTAGGGCCAGACTGAACACCCGAGGCCTCTGAGA
 617 TACAGCCCAGGGCATTCTGCAATGGGAGCACCAATCAGAGAGAAAAGGACACAGGACCCTTCAGGCATGCACCCTAGTGTACACTTGG
 705 GACTGGTAGGTTTTCCTATCCCTGGAGCATTGGCTGTGCTCTGTTTGCGCAAGCTTACAGGAGCATCTCATCTTTGCACATGAAAATT
 793 CTTTTCCCTCACAGGACCCAATGGACCTCTCGCTGATGCACATTTAGGCCTCTTTCTGTTGCAGCGAATCAGAGATACTTACATAAGG
 881 AGATGGGCAAAATGGAGATGGATTGTAGCTGTGGCCCAATGAAATGAAGCCACGGGCCATCCATCACCAGGAAAGCAAATGTGAGTGG
 969 ATGAGGCATGCTGGGCCCCTTGATTTATGCGCAGAGTGCTGGTACCCAGTGCGCGTGTGCAAGCTGGTACACCAGCCAATCAGAGAAC
1057 ATTATATCAGCCAAGGCCCGACCAAGTGGGATGAACCTCTCAGTGGAAGGAGGCTGCCTTCTGGAGAGGAGGGGACTAGCTGGGTGAC
1145 GAGTTTTGCCGGGGGGGTGGGGTGGGGGGGGGTGGGGACGTTCCACAGTGGGACTCCGCGGGTGGTTAGGAGGCACCCTTGTTGTCCTC
1233 GTCGTCATTGTAGTGGGACAGTGGGGTCACAGAATATGAGGGTCCCGGGAAGCCTGTGAGGTGACTTCCGGAGCTTTCTTTTACTGTA
1321 AGAGGCGCCAGAGCTGAGAATGAGAAGCCAGGAATGAGAAAAACTCCCAGTTTGGGGTTGTTTTTCGTTGTTCTTTGGTTTTGTTTGG
1409 TTTTTGTGGTGGTGGTGGTTTTTTTTTTTTTTTTTTTTTGTTGTTGCTGGGTCTTGTTGTTGTTTGTTTGTTTTGTTTGGTTG
1497 TTTTTTTTTGTTTGTTTGTTTGGTTCCTGACTGCACACCTGGATCCTCCTGTGACACTTCCATAAGCAGTATCTGAGCAGGATAGTTA
1585 GTCTGAAGCGGAAGGGGGAAACCCTGGGCACTTTGCACAAACACATGGGCAACGACAGGGCCCAAGCTGGGAATGGGTACCTGATTGG
1673 CAGGCCTCACACCTTTGGGAGTACCTGTAGGCAGCCTCAGGCATGCCTGCACAGTAGTGTTTCTGATAATCAGAGACCTACTCCAGAA
1761 TGTGGGGCCTCTATCTGAGGGGCGGGGGAAATAGCCTCTCAGAGGAATGTATTGAAGTGTGGTCACCTGGGTGAGTGACAGGGAATTA
1849 GCAGAGTCAGTCTTGAGCCTCTGATGGTGAGAAAGCTCTGTCAGCCTGGTGTATCTGGCCTTTTGGGGAGCAAAGTGGGGTCTTGAGG
1937 CAGAAGGGACCAGGTTTAGCTGTGATGAGATGTTTGGAAGACTTTGCACCTGGTAGTAAAATAGACCAGTGAGACCAGTTGATGAAAA
2025 TGGTTCCAGGAGCGTTGCAGGTTTGAAGAGAGTTGGCTGGGACAGTGACTTGGAGAAAAGGAAGTAGTGACTCTCCCCTGTCCTCCAG
```

Figure 3.2.5: **4C Primer Design in *Xlr3b*.** 4C primers within the *Xlr3b* sequence (HindIII library primers, teal; DpnII library primers, green; primary restriction sites, yellow; secondary restriction sites, red; Atrx binding region determined by bisulfite sequencing, grey [96]); Xlr3b Exon 1, blue.

Lastly, I wanted to confirm CTCF enrichment within intron 1 of *Xlr3b*, seen in ChIP-Seq data available on the UCSC Genome Browser (Figure 3.2.6). CTCF has held much interest, as it is the insulator protein that binds to the *H19* ICR. While it is unlikely that an ICR lies within the imprinted gene itself, I was curious to see if there was differential binding at this CTCF site–or any binding at all. The same peak also exists in *Xlr3a*, a biallelically expressed paralog. Figure 3.2.7 shows ChIP-qPCR results with paralog-specific primers in *Xlr3b* and *Xlr3a*. While overall, CTCF enrichment is lower in *Xlr3b* compared to *Xlr3a*, there is a nearly three-fold enrichment of maternal CTCF over paternal at *Xlr3b*, and equal enrichment at *Xlr3a*.

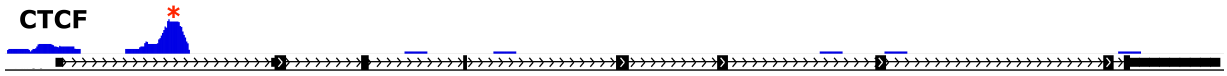


Figure 3.2.6: **CTCF signal across *Xlr3b*.** CTCF pile-up in intron 1 of *Xlr3b*, seen in MEL cell line (red asterisk). Figure adapted from ENCODE.

It is highly improbable that this is the imprinting mark that is causing the differential expression in *Xlr3b*. Rather, like RNA pol II binding, CTCF reflects active expression of the maternal allele, but tells us very little about why the paternal allele does not carry this binding to the same degree.

3.3 Discussion

To date, over a decade of work in the O'Neill lab has been put into elucidating the imprinting mechanism behind X-linked genes, *Xlr3b, Xlr4b*, and *Xlr4c*. At first glance, it appears that these genes defy classic mechanisms of autosomal imprinting. The apparent lack of differential methylation around the locus and evidence for co-

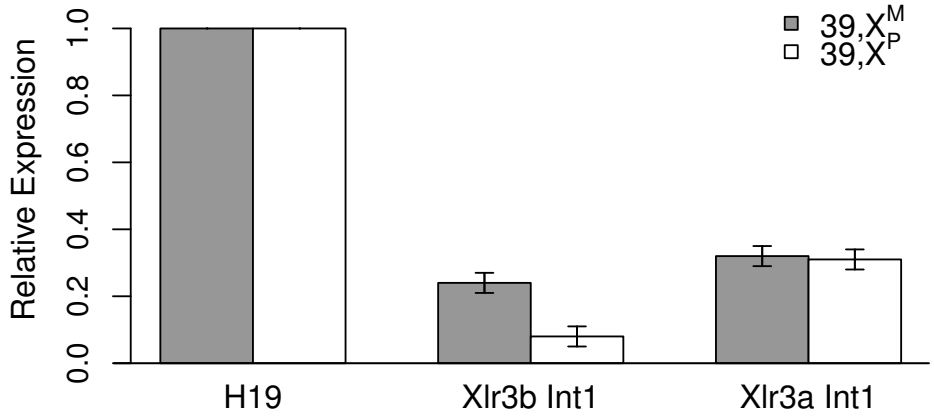


Figure 3.2.7: **CTCF enrichment at *Xlr3a* and *Xlr3b*.** CTCF ChIP-qPCR on neonatal brain, showing slight maternal enrichment of CTCF at *Xlr3b* and equal enrichment at *Xlr3a*, normalized to *H19*. N=3.

transcriptional regulation led to a seemingly novel mechanism to regulate PSGE of this imprinted locus. The first breakthrough came in the form of *Factor 8-associated gene (F8a)* as a potential ICR. A CRISPR deletion of *F8a*, however, did not significantly affect *Xlr3b* expression.

The single-copy gene *F8a* is located within the imprinted locus, between *Xlr4b* and *Xlr4c*. Foley's analysis began with a search for differential methylation of a CpG island located within *F8a* by bisulfite sequencing and a HpaII RT-PCR assay. He saw ~20% methylation in X^M and ~1% methylation in X^P using these techniques in both neonatal brain and fibroblasts. In addition, using ChIP-qPCR, he found enrichment of CTCF occupancy on the paternal allele and enrichment of MeCP2 on the maternal allele. This was suggestive of a model in which CTCF binds to the unmethylated paternal allele, while MeCP2 recruits Atrx to areas of high G4 formation within the imprinted paralogs to allow maternal RNA pol II progression through these gene bodies. Attempts to reproduce this methylation and ChIP data, by Foley and myself,

were not successful.

Reproducibility of data and inconsistency between samples has made this a frustrating journey, to say the least. Tracking down the ICR has been a bit like looking for a needle in a haystack. Another caveat is the inherent repetitive nature of the X-chromosome, which contains a high number of ampliconic gene families. Much existing publicly available data is not informative when it comes to the X-chromosome, at least not when it comes to our region of interest. Furthermore, ENCODE data for the X-chromosome is almost exclusively from males. Due to the challenges in analyzing and interpreting X-chromosome data, the X is often left out of genomic analyses altogether.

With the emergence of greater sequencing technologies, longer read lengths, and increasingly innovative molecular techniques, improved methods for analysis will be a turning point in the search for the molecular mechanism behind *Xlr3/4* PSGE. This will be covered in the following chapter.

Chapter 4

Synthesis and Future Directions

The initial aim of this dissertation was to characterize the nuclear architecture surrounding the imprinted *Xlr3/4* locus, in hopes of elucidating an imprinting mechanism for PSGE of *Xlr3b, Xlr4b,* and *Xlr4c*. Existing studies have shown that imprinted autosomal loci show differences in interaction profiles that reflect previously discovered epigenetic mechanisms for transcriptional regulation. For this X-linked locus, however, mechanisms of regulation are still not understood. Many difficulties encountered during efforts to determine how these genes are paternally silenced, stem from the repetitive nature of the X-chromosome, as the imprinted genes are but three of many paralogs that are nearly identical at the nucleotide level. These efforts thus far have found: a lack of differential methylation surrounding the imprinted locus, and a potential co-transcriptional model of gene regulation. Difficulties in reproducing previous results have led to a new direction for this work.

In recent years, several reports suggest H3K27me3 may be an imprinting mark for methylation-independent genomic imprinting and paternally imprinted XCI. ATRX was also added to the list of genome regulators that include CTCF and MeCP2, that are involved in genomic imprinting and the regulation of PSGE. Limitations of ChIP-Seq methods are tempered by depth of coverage obtained for the X-chromosome, as

the singular X sees lower coverage compared to all other chromosomes. Additionally, even with the increased mapability of longer sequencing reads, the X-chromosome encounters lower alignability compared to the autosomes. These limitations make genome-wide methods costly and sometimes ineffectual. While the ChIP-Seq results presented here do not show differential enrichment at the *Xlr3/4* locus, there are still dramatic global differences between X^M and X^P. Combined with the differences in nuclear organization of the maternal and paternal X-chromosomes, the work presented here certainly raises some broader biological questions about the mouse X-chromosome.

For the imprinted *Xlr* copies, in the context of gene dosage, what is the functional significance of only having one highly expressed allele in females? The function of the *Xlr* genes is still not well understood. The *Xlr* multi-copy gene family encodes multiple proteins similar to *Synaptonemal Complex Protein 3 (Sycp3)*. The most well-studied members of the *Xlr* family are the *Slx* and *Sly* genes in mice [97][98]. These genes are exclusively expressed in post-meiotic spermatids and have contrasting effects in that *Slx* stimulates XY gene expression in spermatids, while *Sly* plays a role in repressing XY genes of post-meiotic sex chromatin (PMSC). While studies of the Xlr3 protein are currently underway, we so far understand that the protein colocalizes with the XY body during male meiotic sex chromosome inactivation (MSCI). A *Xlr3* shRNA knockdown transgenic mouse model knocks down *Xlr3a*, *Xlr3b,* and *Xlr3c* transcripts in spermatocytes leading to spermatogenic defects as well as disruption of MSCI.

In testis, the Xlr3 protein plays a critical role in male meiosis–thus, it is essential that the maternal copy is never silenced. While these genes are highly expressed in testis, the imprinted copies also show imprinted expression in brain and ovaries. In the *Xlr3* knockdown brain, we see a surprising observation of upregulation of the imprinted genes in the transgenic knockdown, which suggests that the Xlr3 protein may be regulating its own imprint (Naveh, unpublished). The function of the Xlr3 protein in brain is still a mystery and it raises several important questions. If the protein has some detrimental effect in the brain, why imprint just one copy of the *Xlr3* paralogs? The segmental duplications of ampliconic gene families, observed on the X-chromosome, results in multiple highly similar copies of *Xlr3* and others. Are some of these backup copies? It is interesting that *Xlr3b* is more highly expressed in brain compared to the biallelic low expression of *Xlr3a*. What about *Xlr4* and *Xlr5*? Do they share a similar function with *Xlr3*–possibly also as a transcriptional regulator? Shioda *et al.* reported that *Atrx* knockdown resulted in decreased methylation at the *Xlr3b* TSS and increased *Xlr3b* expression in various regions of the brain. How Atrx regulates *Xlr3b* expression is still not well understood. But as *Xlr3* upregulation in mice is accompanied by cognitive deficits, this suggests that the imprinted expression may be far more relevant in the brain, such that high expression of even a single active copy could have unfavorable effects.

The results of the work presented in this dissertation suggest that there are broad-scale epigenetic differences between the maternal and paternal X-chromosomes, such that the X^M interactome is enriched with *cis* genes related to intellectual disability

and cognitive impairment. I hypothesize that this might be related to the maternal X susceptibility to ASD, as well as several other neurodevelopmental disorders (ND) that show a gender bias [99]. Further work should be done to validate these results. 3D-FISH can be used to detect the spatial arrangement of various X-linked targeted loci in the nucleus. These *in situ* experiments allow for the simultaneous visualization of differently labeled nuclear targets and would reveal not only the spatial relationship of multiple loci with each other, but also their position within the nucleus itself.

To definitively answer the question of whether or not the maternal and paternal X-chromosomes are organized differently within the cell, a technique such as Hi-C can be employed. Like all 3C-related techniques, Hi-C relies on proximity-based ligation, and requires a high degree of sequencing depth to capture all pairwise interactions between fragments. Hi-C also shares the caveats of 3C-related techniques. Ligation-based detection cannot capture all interactions, and in particular, proximity-based ligation misses many significant interchromosomal interactions.

A newer method to detect nuclear organization solves this problem. This assay, called Split-Pool Recognition of Interactions by Tag Extension (SPRITE), is not based on proximity ligation [100]. Instead, SPRITE relies on barcoding interacting molecules using a split-and-pool approach. This technique is able to detect wider-spaced DNA-DNA interactions. Additionally, SPRITE can also identify DNA-RNA interactions. Preliminary studies of this technique were able to replicate the results of Hi-C experiments as well as *in situ* imaging studies. The resolution of the ex-

periment is determined by bin size. Fine-scale resolution would allow focus on the imprinted *Xlr3/4* locus and large-scale resolution would be able to look at the entire X-chromosome.

Long-read sequencing is currently underway to resolutely answer the long-standing question, is there a DMR that controls the imprinted expression of the *Xlr3/4* locus? Initial searches for allele-specific differential methylation found no DMR. Instead, these experiments found hypermethylation of all CpG islands, with the exception of *F8a*. The difficulties in understanding the epigenetic environment of this locus always go back to the ampliconic nature of these paralogs. This complication makes it difficult not only to map reads from high-throughput sequencing, but also to design primers to capture regions for smaller scale analysis, such as bisulfite sequencing and ChIP-qPCR. Few solutions currently exist that combine long-read sequencing with epigenetic solutions.

For many years, our lab has been focused on determining the mechanism behind X-linked imprinting regulation. Thus far, very little makes sense about what we've found. About half a dozen genomic features seem to act in unison to silence *Xlr3b*. Dnmt1 knockout cells saw an increase in *Xlr3b* expression [39]. HDAC1-deficient cells saw an increase in *Xlr3b* expression [42]. Histone H1 depletion resulted in increased *Xlr3b* expression [43]. 5-aza-dC treated cells also saw increased *Xlr3b* expression (Kasowitz, unpublished). Transcription factors Cux1 and Cux2 have also been shown to be negative regulators of *Xlr3b* and *Xlr4b* expression [101]. Most recently, an *Atrx* knockdown resulted in increased *Xlr3b* expression [96]. It

would seem that multiple genomic features, not limited to histone acetylation, DNA methylation, and chromatin remodeler Atrx, are actively working to silence this gene. Now, we have evidence that the Xlr3 protein, which itself may be a chromatin regulator, may also be negatively regulating its own expression.

An increase in expression of the imprinted *Xlr3/4* alleles upon *Xlr3* knockdown is a curious finding, and begs the question: how does knocking down a transcript lead to its upregulation? And is it only the imprinted paralogs that are upregulated? It would be interesting to see the genome-wide effects of the *Xlr3* knockdown model. RNA-Seq and ATAC-Seq experiments are just a few examples of where we could go with this novel mouse model. Future characterization of chromatin differences in the *Xlr3* knockdown could reveal more about its role as a chromatin remodeler. As Shioda reported, an increase in *Xlr3b* expression in an *Atrx* knockdown mouse model also saw an increase in neuronal pathogenesis and cognitive defects. As our *Xlr3* knockdown also results in increased *Xlr3b* expression, behavioral testing should fully characterize any cognitive deficits that may exist in these mice.

Because the Xlr3 protein is necessary for MSCI in testis, perhaps it is the high expression of these genes in the brain that is detrimental. But the epigenome cannot silence the maternal copy, as it is the only one inherited by males. So in an effort to reduce expression, the paternal copy was silenced. Due to their imprinted status, the imprinted paralogs have been implicated in the cognitive defects associated with Turner syndrome and ASD. The maternal *Xlr3/4* 4C interaction profile showing a greater number of interactions than paternal, follows from its active expression

and imprinted status. But how do we explain other non-imprinted genes showing similar disparities between maternal and paternal interactomes?

From Skuse's study, it would appear that having only a maternal X-chromosome carries a disadvantage, e.g. 45,X^M compared to 45,X^P and 46,XY males compared to 46,XX females [8]. As the X-chromosome contains a disproportionately high number of genes related to intelligence and cognition, perhaps these differences in nuclear architecture between X^M and X^P could be related to the maternal X susceptibility to cognitive and social impairment. The differences in gene interaction profiles on the X offer clues to explain the variability in cognitive and behavioral phenotypes seen in individuals that inherit a single X-chromosome.

The work presented here opens more questions than it answers, and points future directions to the function of the Xlr3 protein in brain, its potential regulation of the imprinted *Xlr3/4* paralogs, and the status of these genes as candidates for ASD. Finally, the perceived differences in nuclear architecture between the maternal and paternal X-chromosomes suggest that there are parent-specific chromatin behaviors that involve a subset of X-linked genes linked to intelligence and cognitive function.

Chapter 5

Materials and Methods

5.1 Animal Breeding and Tissue Collection

Breeding scheme as depicted in Figure 5.1.1. 39,X^M mice were produced by mating *C3H/Paf* males to *C57BL/6J* females, while 39,X^P mice were produced by mating *C3H/InX1h* females to *C57BL/6J* males, yielding X-monosomic females whose single X-chromosome is derived from the *C57* strain. Sex of neonates was determined visually and confirmed by PCR. Genomic DNA was extracted for genotyping by standard Reagant B/Proteinase K digestion of limb or tail. 39,X monosomics were genotyped for *C3H* versus *C57* size polymorphisms on the X-chromosome using *DXMit130* marker primers. Sexing of 39,X^M neonates was confirmed by amplification using Y-chromosome specific *Smcy* primers, while sexing of 39,X^P neonates was confirmed by size polymorphism of *DXMit130*. All mouse protocols are approved by the University of Connecticut Institutional Animal Care and Use Committee.

5.2 *Xlr3b* Expression Analysis

RNA was extracted from samples using NucleoSpin RNA (Macherey-Nagel), as described by manufacturer standard protocols. RNA was reverse transcribed with

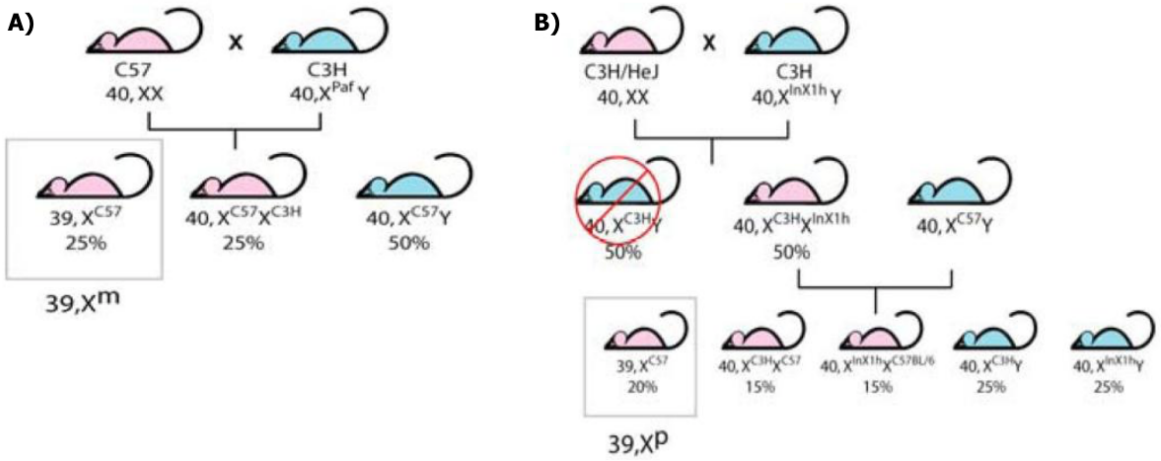


Figure 5.1.1: **Breeding scheme for the generation of X-monosomic mice.** Courtesy of Michael O'Neill.

qScript cDNA Supermix (Quanta) in a volume of 10 - 20 μL, at a concentration of 50 ng/μL. Incubation conditions: 25°C for 5 min, 42°C for 30 min, 85°C for 5 min.

RT-PCR experiments were carried out in triplicate 15 μL reactions with iTaq Universal SYBR Green Supermix (BioRad). *Xlr3b* imprinted expression was confirmed using *Xlr3b* Imp primers, with *ActB* primers acting as a control. The program was carried out on a BioRad C1000 Thermal Cycler with the following cycling conditions: initial denaturation 95°C for 3 min, 45 cycle amplification at 95°C for 10 sec, 58°C for 20 sec, and a plate read, followed by melt curve analysis 55°C - 95°C, in 0.5°C increments. Melt curves and CT values were analyzed with CFX Manager Software.

5.3 Chromosome Conformation Capture (3C)

5.3.1 Mouse Neonatal Brain Tissue

Adapted from Hagege et al [102]. Neonatal brain was weighed, the tissue minced with a blade, and homogenized with 18G and 21G needles in a 1 mL solution of filter

purified 0.125% collagenase (type I) in 10% FBS/PBS. Tissue was incubated at 37°C for 40 minutes, shaking, then strained through a 70 μm cell strainer. Volume was increased to 5 mL with 10% FBS/PBS. 5 mL of 3% formaldehyde in 10% FBS/PBS was added to a final concentration of 1.5% and incubated for ten minutes at room temperature, rocking. 2.5M glycine was added to a final concentration of 0.125M and incubated for five minutes at room temperature, rocking, then for an additional fifteen minutes on ice. The sample was centrifuged at 800xg, 4°C for 8 minutes, the supernatant removed, followed by a cold PBS wash, with another spin to remove the supernatant.

Pellet was resuspended in 1 mL of cold lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 0.3% NP-40, 1X complete protease inhibitors) and incubated for 20 minutes on ice, inverting every 5 minutes. Cells were centrifuged at 2000xg, 4°C for 5 minutes, and the supernatant removed. Pellet was washed by resuspending in 1X restriction enzyme buffer, with another spin to remove the supernatant.

Pellet was resuspended in 450 μL of dH_2O. A 5 μL aliquot was taken as an undigested (UND) control. 60 μL of 10X restriction enzyme buffer was added, followed by 15 μL of 10% SDS to a final concentration of 0.03%. Sample was incubated at 37°C for 1 hour, shaking. 54 μL of 20% TritonX-100 was added to a final concentration of 1.8%. Sample was incubated at 37°C for 1 hour, shaking. 30 μL (300U) of restriction enzyme was added. Sample was incubated at 37°C for 1 hour, shaking. An additional 10 μL (100U) of restriction enzyme was added. Sample was incubated at 37°C overnight, shaking. A 5 μL aliquot was taken as a digested (DIG)

control. Digestion efficiency check with UND and DIG samples was performed with Proteinase K reverse crosslinking, followed by a single phenol-chloroform extraction. After RNase A incubation, the sample was run on a 0.7% agarose gel.

Restriction enzyme was heat inactivated by incubating the sample at 65°C for 20 minutes. Sample was placed immediately on ice. Sample was transferred to a 50 mL conical tube containing 5.7 mL of dH_2O, 750 uL of 10X T4 DNA Ligase buffer, 75 uL of 10 mg/mL BSA, 375 μL of 20% TritonX-100, and 10 uL (4000CEU) T4 DNA Ligase. Sample was incubated at 16°C overnight.

50 uL of 10 mg/mL Proteinase K was added and the sample was incubated at 65°C overnight to reverse crosslinks. 50 uL of 1 mg/mL RNase A was added and the sample was incubated at 37°C for 45 minutes. After phenol-chloroform extraction, the volume of the sample was brought up to 10 mL with dH_2O. Ethanol precipitation was done including 10 uL of 5 mg/mL glycogen, followed by a 70% ethanol wash. Pellet was briefly air dried and resuspended in 150 uL of 10mM Tris-HCl at 37°C. 2 uL of 3C library was run on a 0.8% agarose gel, and successful first round template confirmed by PCR of chimeric ligation products (HindIII: Gapdh5/6 (422 bp), NcoI: ActB1/2 (166bp), DpnII: Gapdh6/7 (184bp)). Libraries were stored at -20°C.

5.3.2 Mouse Fibroblast Cultured Cells

Fibroblast cells may exhibit extensive nuclei aggregation, resulting in poor digestion efficiencies [103]. Fibroblast libraries used a cell scraper to collect adherent cells and included a dounce homogenization step with pestle A after cell lysis.

5.4 3C-Carbon Copy (5C-Seq)

5.4.1 3C Control Library Generation

A 3C control library was made using five BAC clones covering two genomic regions of interest (Supplementary Table A.2.6). The BACs were purified with a Plasmid Midi Kit (Qiagen) and mixed in equimolar ratios. 12 ug of BAC DNA was digested with NcoI overnight. After phenol-chloroform extraction, ethanol precipitation, and a 70% ethanol wash, a 200 uL ligation reaction was performed at 16°C overnight. The ligase was heat inactivated by incubating the sample at 65°C for 15 minutes. After two phenol-chloroform extractions and an ethanol precipitation, the volume of the sample was brought up to 500 uL with TE buffer. The sample was purified with a 30K Ultra Centrifugal Filter Unit (Amicon) and the volume adjusted to 100 uL. After RNase A incubation, DIG, UND, and control library samples were run on a 0.8% agarose gel. The control library was stored at -20°C.

5.4.2 5C Primers and Library Generation

Adapted from Ferraiuolo et al [104]. A NcoI 3C library was generated from mouse primary fibroblast cells. The my5C.primers module [105] was used to design 184 alternating forward and reverse 5C primers (Supplementary Table A.1.4) spanning two genomic regions of interest: the imprinted *Xlr* locus and the *Tktl1* locus (chrX:70,225,339 - 70,690,339, chrX:71,352,340 - 71,666,339). All 5C primers were resuspended at a 50 uM concentration. Forward primers were pooled. Reverse primers were pooled and phosphorylated in a PNK reaction. Forward and phos-

phorylated reverse primers were pooled and diluted such that individual primers were at a concentration of 1.7 fmol/uL.

Five reactions with 750 ng 3C template (5 ng for BAC), in addition to a no primer control, a no ligase control, and a no template control were set up per 5C library. Primers were annealed at 48°C overnight. Ligation mix was added the next day and samples were incubated at 48°C for one hour, and then at 65°C for 10 minutes. Libraries were amplified for 24 cycles using universal T7/T3 primers. Libraries were pooled and purified using MinElute PCR Purification Kit (Qiagen). Quality control was performed with serial dilutions and nested PCR.

5.4.3 Illumina MiSeq Sequencing

The TruSeq ChIP Sample Preparation Kit was used to ligate adapters to pool multiple experiments within one sequencing lane. Libraries were visualized with BioAnalyzer 2100 High Sensitivity DNA Assay (Agilent) and quantified by qPCR. The MiSeq Reagent v3 kit (150 cycles) generates approximately 25 million 2x75 paired end sequencing reads.

5.4.4 Data Analysis

FASTQ files were uploaded to Galaxy (https://usegalaxy.org) for secondary sequence quality control (QC). Sequences were filtered to remove reads shorter than 70 bp and converted to FASTA format. 5C reads were mapped to a pseudo-genome of all possible forward and reverse primer pair sequences (92 x 92 = 8464 total possible unique pairwise interactions). Files of interaction counts were generated for all

libraries. The my5C.heatmap module [105] was used to visualize the data.

To normalize for primer efficiency, interaction frequencies were calculated by dividing the amount of a 5C ligation product by the amount of the same ligation product in the control 5C library. This ratio is a direct measure for the frequency of interaction. 5C control library deficiencies (i.e. overlapping regions, uncovered regions) along with primer design in a repetitive region of the genome contributed to an overall unsuitable method.

5.5 Circular 3C (4C-Seq)

5.5.1 4C Primers and Library Generation

Adapted from Splinter et al [106]. Secondary digestion of DpnII 3C libraries was performed with 50U of Csp6I in a total volume of 500 uL at 37°C overnight. A 5 μL aliquot was run on a 0.7% agarose gel as a digestion efficiency check. Restriction enzyme was heat inactivated by incubating the sample at 65°C for 25 minutes. Sample was transferred to a 50 mL conical tube for a 12 mL ligation reaction with 10 uL (4000CEU) T4 DNA Ligase. Sample was incubated at 16°C overnight. Ethanol precipitation was done including 10 uL of 5 mg/mL glycogen, followed by a 70% ethanol wash. Pellet was briefly air dried and resuspended in 150 uL of 10mM Tris-HCl at 37°C. Samples were purified with a PCR Purification Kit and eluted into 50 uL of 10mM Tris-HCl. 1 uL of the final library was run on a 2% agarose gel. Libraries were stored at -20°C.

4C primers were designed such that the reading primer hybridizes adjacent to

and overlaps the primary restriction enzyme site. This fixed position ensures the analysis of primary ligation products only. Non-reading primers were designed near a secondary restriction enzyme site, such that the viewpoint fragment is large enough to allow efficient crosslinking to other DNA fragments, and the fragment end is large enough to allow efficient circularization. Primers were designed inverse and were checked against the mouse genome to maximize uniqueness. Primers were tested by 4C template titration to assess the quality of the 4C library as well as the functionality of the primers, using Expand Long Template PCR System (Roche).

5.5.2 Illumina NextSeq500 Sequencing

For sequencing, primers were made with 5' overhangs composed of adapter sequences necessary for Illumina single read sequencing on the NextSeq 500 platform, added to the gene-specific sequences (Supplementary Table A.1.5). The Amplicon PCR was carried out with eight 50 uL reactions per library, amplifying 200 ng of 4C template per reaction for a total of 1.6 ug per library, with the following cycling conditions: initial denaturation 94°C for 2 min, 24 cycle amplification at 94°C for 10 sec, 56°C for 1 min, 68°C for 3 min, and a final extension 68°C for 5 min. Reactions were pooled and purified with AMPure XP beads.

The Nextera XT Index Kit attaches dual indices and Illumina sequencing adapters to pool multiple experiments within one sequencing lane. This Index PCR was carried out with 8 cycles of amplification. Libraries were visualized with TapeStation 2200 D1000 High Sensitivity DNA Assay (Agilent) and quantified with Qubit dsDNA

HS Assay Kit (Thermo). The NextSeq 500 High Throughput v2 kit (75 cycles) generates approximately 400 million single end sequencing reads. The 75-mer sequencing reads are composed of the 4C primer sequence, specific for a given viewpoint, followed by 55 nucleotides that identify a captured sequence.

5.5.3 Data Analysis

4C-ker, a publicly available pipeline, was used to analyze 4C data. 4C-ker is a Hidden-Markov Model-based pipeline, available as a R package, that identifies regions throughout the genome that interact with the bait locus [107]. 4C sequencing reads are mapped to a reduced genome consisting of unique sequences adjacent to all primary restriction enzyme sites in the genome. 4C-ker then uses overlapping windows of adaptive sizes to identify high-interacting domains in near-bait regions, and across far-*cis* and *trans* chromosomes. In addition, 4C-ker applies DESeq2 differential analysis to quantitatively compare interactions in multiple 4C-seq datasets across different experimental conditions and sample types.

w4cseq is an additional pipeline used for comparison purposes [64].

***cis* analysis**: *Xlr3b*, HindIII, k = 10, *Xlr3b*, DpnII, k = 15, *Xlr4c*, DpnII, k = 15, ***trans* analysis**: *Xlr3b*, HindIII, k = 18, **nearBait analysis**: *Xlr3b*, DpnII, k = 5, *Xlr4c*, DpnII, k = 5, **Replication timing**: *Xlr4c*, DpnII, k = 15 (*cis*), **ChIP-Seq enrichment**: *Xlr4c*, DpnII, $X^{M}133$, $X^{P}268$, size inter = 500, size intra = 100, window intra = 3000 (distal sites), **Repeat Masker**: *Xlr4c*, DpnII, k = 15 (*cis*), **GREAT**: *Xlr3b*, DpnII, k = 15, *Xlr4c*, DpnII, k = 10.

5.6 Chromatin Immunoprecipitation (ChIP-Seq)

5.6.1 ChIP Library Generation

ChIP was performed using the MAGnify Chromatin Immunoprecipitation System ChIP Kit (Invitrogen). Neonatal brain was weighed, the tissue minced with a blade, and homogenized with 18G and 21G needles in ice cold 1X PBS. 37% formaldehyde was added to a final concentration of 1% and cells were incubated for ten minutes at room temperature. Formaldehyde was quenched with glycine for another ten minutes at room temperature. The sample was centrifuged at 300xg, 4°C for 10 minutes, the supernatant removed, followed by two cold PBS washes. The cells were then lysed in protease inhibitors for twenty minutes on ice. Chromatin was sheared to 100 - 300 base pair fragments using the Covaris S2, with the following conditions: Duty cycle = 5%, Intensity = 2, Cycles per burst = 200, Cycle time = 60 seconds, Cycles = 10, Temperature = 4°C, Power mode = frequency sweeping, Degassing mode = continuous. Sonicated chromatin was digested with proteinase K and visualized on a 1.5% agarose gel to confirm appropriate size range. Chromatin was stored at -80°C.

Antibodies were coupled to Protein A/G Dynabeads for a minimum of 4 hours at 4°C, rotating end over end (Supplementary Table A.3.7). 10 to 20 ug of sheared chromatin, approximately 200,000 cells, was diluted and bound to antibody-Dynabeads overnight at 4°C, rotating end over end. The next day, the bound chromatin was washed, crosslinks reversed, and DNA purified in a volume of 30 uL of elution

buffer. The DNA was tested by qRT-PCR with positive and negative control primers (Supplementary Table A.1.3) and stored at -20°C.

5.6.2 Illumina NextSeq500 Sequencing

The TruSeq ChIP Sample Preparation Kit was used to ligate adapters to pool multiple experiments within one sequencing lane. Libraries were visualized with TapeStation 2200 D1000 High Sensitivity DNA Assay (Agilent) and quantified with Qubit dsDNA HS Assay Kit (Thermo). The NextSeq 500 High Output v2 kit (300 cycles) generates approximately 400 million paired end sequencing reads.

5.6.3 Data Analysis

The results from multiple sequencing runs were pooled for data analysis. ChIP-Seq reads were mapped to the mm9 genome using Bowtie2 and converted to bedgraph format for visualization. Peak calling was done using MACS2, standard peaks for Atrx and broad peaks for H3K27me3. ChIPseeker, a publicly available R package, was used for annotation, comparison, and visualization [108].

5.7 Assay for Transposase-Accessible Chromatin (ATAC-Seq)

ATAC libraries made from mouse neonatal brain tissue and sequenced on the Illumina NextSeq500, data provided by Glenn Milton. ATAC-Seq reads were mapped to the mm9 genome using Bowtie2 and converted to bedgraph format for visualization. Peak calling was done using MACS2. ChIPseeker, a publicly available R package, was used for annotation, comparison, and visualization [108].

Appendix A

A.1 Primers

Table A.1.3: General Primers

Primer	Sequence	Temp (°C)	Size (bp)
MIT130 F	TTCATATCGCCCCAACCTAC	56	157/169
MIT130 R	TATTTTGAAACCTCTGCCATTT	56	157/169
Smcy F	TGAACTGCCTGCTATGCTAC	60	283
Smcy R	GCCTCAGATTCCAATGCTC	60	283
Xlr3b Imp F	AGGCTGCCTTGTGGAGAG	59	95
Xlr3b Imp R	TGTCAGTGGCCTTCCTTTTT	59	95
ActB RT F	ACACCCGCCACCAGTTCG	59	123
ActB RT R	CGATGGAGGGGAATACAGCC	59	123
Gapdh 5	GGCCTGTTTAAGGGAAAGAAGT	58	variable
Gapdh 6	CCCACCCTAGAAAGTCCAAAG	58	variable
Gapdh 7	CGTGAGTGGAGTCATACTGGAA	58	variable
ActB NcoI 1	CGATGGAGGGGAATACAGC	58	166

Continued on next page

Table A.1.3 – *Continued from previous page*

Primer	Sequence	Temp (°C)	Size (bp)
ActB NcoI 2	CTAGGCGTAAAGTTGGCTGTG	58	166
T7	TAATACGACTCACTATAGCC	60	130
T3	TATTAACCCTCACTAAAGGGA	60	130
Xlr3b Pro F	GACTAGTAACCCCTGGAGACTCAT	58	356
Xlr3b Pro R	CATTGGGTCCTGTGAGGG	58	356
Xlr3b Int3/Ex4 F	TTGTCCTGCAGTCCATCAGAC	58	381
Xlr3b Int3/Ex4 R	CCTTCGAATTCTTGGACCAAA	58	381
Xlr3b Int7 F	CGTTCCATCTTTAAGCAGACTAGTT	58	231
Xlr3b Int7 R	ACATGCCTAGCAAGAACTTACAGAC	58	231
Xlr3b Ex9/3′UTR F	GAATACAGCTTAATTCACCATGG	58	171
Xlr3b Ex9/3′UTR R	CCATACAGGAAATCACAGGACG	58	171
Xlr3b TSS F	CCAAGTGGGATGAACCTCTGA	58	142
Xlr3b TSS R	GTGCCTCCTAACCACCCG	58	142
Xlr3a Int1 F (CTCF)	AGTTGGCTGGGACAGTGACA	58	229
Xlr3a Int1 R (CTCF)	CAACATTCAACACAGATTTTAGGAG	58	229
Xlr3b Int1 F (CTCF)	AGTTGGCTGGGACAGTGACT	58	229
Xlr3b Int1 R (CTCF)	CAACATTCAACACAGATTTTAGGAA	58	229
H19 F (CTCF)	AAGGGAACCATTCCAGAGGT	58	212
H19 R (CTCF)	CAGCCAGTGTGGCTCACTAT	58	212

Continued on next page

Table A.1.3 – *Continued from previous page*

Primer	Sequence	Temp (°C)	Size (bp)
Dhrsx F (Atrx)	CTCCTATCCCCACATCCTGACT	58	303
Dhrsx R (Atrx)	GAGGTCACACAGGGTAGAAAGG	58	303
Hoxc10 F (H3K27me3)	TTGAGCAGAGGCCTAAGGAA	58	124
Hoxc10 R (H3K27me3)	GCACTTCCATGTCTCGGTTT	58	124
Pax2 F (H3K27me3)	CAAGTTCCCTGGCTCTCCTA	58	94
Pax2 R (H3K27me3)	CAGACGCCTGGTTTGGAAT	58	94

Table A.1.4: 5C Primers

Primer	Sequence
Xlr3b-FOR-2	TAATACGACTCACTATAGCCTTCATGTCCGACGGTGGCTTCCTAGTACAGGGCCCATTAATACCA
Xlr3b-FOR-4	TAATACGACTCACTATAGCCCTCCCGGAACCCTCTGTTGACCTACCACCTACTCCCTCCCACCCA
Xlr3b-FOR-6	TAATACGACTCACTATAGCCCGACCCTCTTTCTTCTCGTATTACACTACAGACTTTGGACACCCA
Xlr3b-FOR-8	TAATACGACTCACTATAGCCAAGAACTGATGGGTCATGTTTTCATATTCTCCAGGTGAGCAGCCA
Xlr3b-FOR-10	TAATACGACTCACTATAGCCTACAGGTAACAGGCTCTGAATAGTTCTTAGAGCAGTGCTTCCCCA
Xlr3b-FOR-12	TAATACGACTCACTATAGCCAGCGTCGGTCCCACTGAGCCTCCTCAATAATCTACTCTCCGCCCA
Xlr3b-FOR-14	TAATACGACTCACTATAGCCTAATAGTTCTTCTGGGAACTAAGAATGCACAATGTCCTAGAGCCA
Xlr3b-FOR-17	TAATACGACTCACTATAGCCGTAAAAGTAAGTTTGCACCACAAAGGGATAAAACATCAGTGCCCA
Xlr3b-FOR-20	TAATACGACTCACTATAGCCCTCGCCGGTCCTCCCTTGGGAAGGAATTAGTATATGCTACTCCCA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Xlr3b-FOR-22	TAATACGACTCACTATAGCCTCTGAATGACCGATTGCTGGACAGTCATCAACGTAGGTTCCTCCA
Xlr3b-FOR-24	TAATACGACTCACTATAGCCACCTCGACCGAGCCCCAGGGCAAGCCTATAAAGAGGCTGGGGCCA
Xlr3b-FOR-28	TAATACGACTCACTATAGCCAGTGTGTATCAGGCAAGTTGAAGTCATCAGCCTTGTCAACATCCA
Xlr3b-FOR-31	TAATACGACTCACTATAGCCCTTTGTGAAACGATACTAGTTAATTACTATGGGGCCAGGGAACCA
Xlr3b-FOR-34	TAATACGACTCACTATAGCCGGATCGATTTGGCTATTTACTATTTCTAGAACCAGCCTCTAGCCA
Xlr3b-FOR-38	TAATACGACTCACTATAGCCATATTTAAGGAACAATACATACATGCTTGACTTCTAAAGAATCCA
Xlr3b-FOR-41	TAATACGACTCACTATAGCCCGAGGGGCCTCAAGGACTCGAGCGTCTGGAAGCCAGCAGGAACCA
Xlr3b-FOR-45	TAATACGACTCACTATAGCCAGACAAGTGACTGGGTTTTCAGAACATCCTTAATTTTGACCTCCA
Xlr3b-FOR-51	TAATACGACTCACTATAGCCGGGGTTCCAAAGGGTCCGGAAGGAGCCATTCATTTGCTGTGTCCA
Xlr3b-FOR-55	TAATACGACTCACTATAGCCGGGTCAGGGTGTCTTGGGGAGAATGTCTCGAGGGTGGGCTGGCCA
Xlr3b-FOR-57	TAATACGACTCACTATAGCCCTCAGTGGGACAGGACTAGTCGTGTGCCCCAGAGTCCAGTCGCCA
Xlr3b-FOR-62	TAATACGACTCACTATAGCCCGAACCCACCACATCATCAGCATAAAGGAGCCCAGGAAGGAGCCA
Xlr3b-FOR-65	TAATACGACTCACTATAGCCCCATTTCGGGAAGGCCTTCGTAGTACTCCCCTCCCCGGCATGCCA
Xlr3b-FOR-67	TAATACGACTCACTATAGCCCGTGTCCCTTGTACCCGGCAGTCCGATACTCATCAGGCTATTCCA
Xlr3b-FOR-69	TAATACGACTCACTATAGCCCGTACACTAACGTAGCTCTGTCTTCTCTGTGAAAACTGCATTCCA
Xlr3b-FOR-71	TAATACGACTCACTATAGCCTGGTCCGAAAGCCCCAAACCTCATGTCTCCACCTCACCCTTGCCA
Xlr3b-FOR-81	TAATACGACTCACTATAGCCGAGTAAAACGAGGTCGTGTTCAGCCAGGATACTCAGCCACACCCA
Xlr3b-FOR-83	TAATACGACTCACTATAGCCGTGACTAAGAGGTGATTCTGAACATGAATACAGCTTAATTCACCA
Xlr3b-FOR-85	TAATACGACTCACTATAGCCCAGGTACTCTCGTCTCAACAACACAATCTCAATACCCCTTGGCCA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Xlr3b-FOR-88	TAATACGACTCACTATAGCCCATATACTAACTAAATAAGCTGTAAACTAGCAAAGGCTCCCACCA
Xlr3b-FOR-90	TAATACGACTCACTATAGCCACGGGTCTGGGACCTCTTCATGAGCTGGGAGGCTTTCGATGGCCA
Xlr3b-FOR-95	TAATACGACTCACTATAGCCACCAACACAAGATCATTGTGGCAAACCCTGCCCCCAAGGTCTCCA
Xlr3b-FOR-106	TAATACGACTCACTATAGCCGACGAACGTTACGAGCTCCATCACTTAAGGAAGGGTTACCTCCCA
Xlr3b-FOR-108	TAATACGACTCACTATAGCCCTCAGTGGGACAGGACTAGTCCACAGCCCCAGAGTTCAGTCACCA
Xlr3b-FOR-113	TAATACGACTCACTATAGCCACTACATTCAGATCTTATCAGACACGCCTGTGTGAGCTGGGCCCA
Xlr3b-FOR-116	TAATACGACTCACTATAGCCAGTATTGTTTTTATCTATTCTCTGTGATAAATTTTTACTTTCCCA
Xlr3b-FOR-119	TAATACGACTCACTATAGCCATCTTGGTTAGCAAACATCTTTGTGTCAGAGGCTTGAGAAATCCA
Xlr3b-FOR-121	TAATACGACTCACTATAGCCGGGCTCTGGGGACGGGGTGGACTTGCCCCTTCCAAAGCTTGCCCA
Xlr3b-FOR-124	TAATACGACTCACTATAGCCCGTGAAAACTCATCGTCGGAGATTTCACAGCACAAGATGGGGCCA
Xlr3b-FOR-126	TAATACGACTCACTATAGCCAAGTGCCAGTATTAGGTAGTAACTCGGAAGCCGGGGCAAGAACCA
Xlr3b-FOR-128	TAATACGACTCACTATAGCCAAATATCGGCAGATCTTTTGTAGTATGGGCTTTAAGAATAGACCA
Xlr3b-FOR-132	TAATACGACTCACTATAGCCCGACAAAGGTAAAGGATGCTGTTGTGGCTAGAGTTGCAATGACCA
Xlr3b-FOR-135	TAATACGACTCACTATAGCCTACGTGACGAAGTGGTCAGGCCCCAAAGTGATGAGTAATCAACCA
Xlr3b-FOR-137	TAATACGACTCACTATAGCCTCTAAGTCTTCGGTGAAGACCTCCCTCTACTTCAAATCCCCACCA
Xlr3b-FOR-139	TAATACGACTCACTATAGCCTACGATCCAGTGAAACCTATGAGGCAGATCAGAGTGACCATCCCA
Xlr3b-FOR-141	TAATACGACTCACTATAGCCGGGCTCTGGGGACGGGGTGGAGAAACCCCTTCCAAAGCTTGCCCA
Xlr3b-FOR-145	TAATACGACTCACTATAGCCGTCAGTGTCTTAGTGCTATATTGCTGTGAAGAAACACTATATCCA
Xlr3b-FOR-149	TAATACGACTCACTATAGCCCTTCCCTAAGGTCGCGTTCGAGGAGGAGCATCTGCGGGACCTCCA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Xlr3b-FOR-151	TAATACGACTCACTATAGCCCTGTTCGGTGTTTTACCCCAAGGTTGCAGAAACTACTAGGAGCCA
Xlr3b-FOR-153	TAATACGACTCACTATAGCCCCTCTACACTCTACAGAGTGTGACCAGAGGCTAGAAGGTTATCCA
Xlr3b-REV-1	TGGGATCCCTAAGCAGACAGGAGTCCACCCAACCCTCTGGACCGTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-3	TGGTTTTCCGGAGTTTTAGTACAGTTGGACGCGTTCCTCCGAGGATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-5	TGGGAAAATAAGAACTTCATAAGCTTCCTACTGCAGCAGAGGTCATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-7	TGGGTAAGAGAAAGGCCCAGAGGCGTGTGACTCAAGTCTCGGGTCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-9	TGGGTATGCCCATAACAAACACCATCACCAGGCAAAGACGAGTAGTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-11	TGGAGTTTCTCTTGGTGATTCTGGTCCTAGGAGACCCACTATCTTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-13	TGGCTTTCATACAGAGGGCAAAACCTCCGTAACCAACATAGACGATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-15	TGGTTTTTTACAAAGCTTGAGTGTATCCTCCAGAGAGAGTACACATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-19	TGGAGCTCACCCTTTCTTCTAGTTACTATCAAGGGGTGGGTACTATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-21	TGGTAGCAAATAATATCACATTTCTTTACCATGAGGCTTAGGGTATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-23	TGGAGGATGTAAGTGACGAGGAGCCAAGTTAAACCCAGTTGACCTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-27	TGGTGTTTTTCACTCCACCTTGTTCAGTCTACGAGGACTGACCAATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-29	TGGCATCACCTGTTGTCCATAAAAGGCGAATTGGATACACCATTGTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-32	TGGTAGTCAGGTATGATTTGTATCAGGACCTGGCATTTAATTGAATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-35	TGGAATAAACCTGAAATCCTATCACAGAGTGGTTTATTATGCAGTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-39	TGGGACCTGCATGCTTTGCTATCCTTGCCCGGTGACCGATCGAGTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-42	TGGAATCCCTCTGCTATTTTTCCCTTCCCTTTCCGGGTTCTCGATTCCCTTTAGTGAGGGTTAATA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Xlr3b-REV-48	TGGCAGATTTTGATGGAAAGAGACAGACAGTCACACCAAAATTCCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-54	TGGGAGGTAACCTTTCCTTAAGTGATGGCACTCGTGACGTTCGTCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-56	TGGACCTGCACAGGATAGAAAGCAGTACAGCTCCTCACGTGCACCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-60	TGGGGTGGTCATGGATGCAGACTCGAGTCTGTGCAGGTCTATGGTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-64	TGGTCCTCCACGTTTTCTGCCCGTTCACTGGACCTTTCGAACAAATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-66	TGGGGCCTATGATAGATACCAGGTCGCAGACGATTCCTGTATGTGTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-68	TGGCAACTTCCAGCGATGAAAACAGAGGTAAAGGTATCCCTTCTGTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-70	TGGCTATGTCTTGATCTTGCTGTTTGGAAAAAAAAAAGGTCTAAATTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-80	TGGCTGCAAAGAAGCAGGACCCTCTACACCGGTGACTCCTGGGGATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-82	TGGCACTGAAAGATGAGGTCATGATCCTTGACTTAGAAAGAGACATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-84	TGGTCCCCCTGCATCGACCACGGTTGTAGAGAACGACACCTAAGATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-87	TGGGAAGAACTGCTTTCCCCACAAGTTTCTTCCCGGTCTAACAAATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-89	TGGAAGAGCGCCTGCTGGCACGCGCGGTGTCGGTCGACCGTGGTCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-94	TGGCAAGCCATAATCAAAAGTAAGAACTCCAGGAAACGGTCCTCATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-96	TGGAGGTCAAAATTAAGGATGTTCTGAAAACCCACTCACTTCTCTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-107	TGGCCCGCCCACCCTGGAGACATTCTCCCCAAGAAACCCTGACACTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-112	TGGCAGGCACTCAGAATATGAAAAATGGGAGGTTCAAGTGAACGTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-114	TGGCCTCCACAAACCCTAAATTCTTAGCCACCCCTTCTTTCCTAATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-118	TGGATGGAAGTGTGAAATTGGATTCCTCAACATAAGGGGGATTAATCCCTTTAGTGAGGGTTAATA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Xlr3b-REV-120	TGGCCACTAGTGGAGTCTCAAGAAGAGGCGGGGTCTCACACGTCTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-123	TGGGCTTTGGGCAGCAGGCATCACCCTCATACGTACACACATTTTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-125	TGGGCCTTCCAGCAGCACAGGTCGAAATGGTGGAAGTCTGTCCGATCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-127	TGGCTATGTTTAGAAGACATTTTCTCAGTTGATGGTTGATGAACCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-130	TGGTGAAGACAATGGAAAGAGCTGTTAAAGAAAGCTTCCTACTTTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-134	TGGCAACAGCAGACTACTCATCTGACATGCCTCTACTTTAGTGTTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-136	TGGGAGAGCAGACCATCCCAAGGGTCTACTATATTTTCACTTTCTTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-138	TGGCACTGATCTTGAGCTAATTTTCTCAATTACCTGTCGAAATACTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-140	TGGAGAAGGGGGCCATCCAGGGACGTTACCACCTAGTCCGCTCCCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-142	TGGTGGGGTACAGCGAGGGCCACGACTCCCGGGACGGACTGTGTCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-147	TGGTCTCCTAGTAAAGGACAGTTCCTGTGAGGTCCTTCCCGTTCCTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-150	TGGAAATGATTATTTAAGAACACATGACAGTCTGCACAGGTGGAGTCCCTTTAGTGAGGGTTAATA
Xlr3b-REV-152	TGGATTATAGAGCTCAAGCATACCCTCTCCCGTGAGAGAAAGATATCCCTTTAGTGAGGGTTAATA
Tktl1-FOR-6	TAATACGACTCACTATAGCCAAAAAAAAAAGGCTCGAAAGCCCTGGCCTTCTTTCACATCCACCA
Tktl1-FOR-9	TAATACGACTCACTATAGCCTTACACGGTTCAGAGACAGGTTCCAGAGACAGTTTTCTACAGCCA
Tktl1-FOR-11	TAATACGACTCACTATAGCCGGTTTCGAAGGAGGTACTGGCATAAAGCTGCCAGCTCCACCTCCA
Tktl1-FOR-15	TAATACGACTCACTATAGCCATCACGACCGGGAGGCTTAGACTGGATTATTTCATTTCACTTCCA
Tktl1-FOR-17	TAATACGACTCACTATAGCCAAAGGACTAACCTTTTACAGCTGTCTGTTATCAGTAAGCCTTCCA
Tktl1-FOR-20	TAATACGACTCACTATAGCCCCTCGGATTGTCCCTTGGAAAGCAGATACTGCTGTTGTTAACCCA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Tktl1-FOR-22	TAATACGACTCACTATAGCCTCACGTACACGAGATCAGAATGACTTGTCAGTATTTCTCCACCCA
Tktl1-FOR-28	TAATACGACTCACTATAGCCGTTTAAGGAGCTGGTGATCGTAAAAGGCCTGGGAAGGCCTTACCA
Tktl1-FOR-34	TAATACGACTCACTATAGCCACGGTGAGATCAAAGGTTATTAAGGTAGACCCCAGGCTTGACCCA
Tktl1-FOR-36	TAATACGACTCACTATAGCCAATCGACCTTCCACCAAGGCAGAAAGACAGTGGGAAAAGAGACCA
Tktl1-FOR-38	TAATACGACTCACTATAGCCTGAAATCTCAAAAGTAGGGAATGGAGGCGTAGATAAAGTAGGCCA
Tktl1-FOR-40	TAATACGACTCACTATAGCCGTTATCAGTTAAGAGAAGGACTTCTATCTCAACCAGACAAAGCCA
Tktl1-FOR-45	TAATACGACTCACTATAGCCTCTTTCTTGTATGACCGAGATAAAGGGAGGACCAGGAGGCAACCA
Tktl1-FOR-47	TAATACGACTCACTATAGCCCACTTCGTTTTCTACAAGTACAAACTGTGCTACTCAGAGGTCCCA
Tktl1-FOR-49	TAATACGACTCACTATAGCCTCCCCGGTGGTATGAACTACCTGATGAGGTAGCTGCCAGGTGCCA
Tktl1-FOR-51	TAATACGACTCACTATAGCCGTTTCCCGGAGTCCGACCAGTGATCCAGCCCATTAACGCCCACCA
Tktl1-FOR-53	TAATACGACTCACTATAGCCCGGGAACGTGTGTCCGTGTGGGTTCCCCCGTGCCAAGCAACACCA
Tktl1-FOR-55	TAATACGACTCACTATAGCCTTGAACAACCCAAACAACCAGGTACCACTTTGGATGCCTGGTCCA
Tktl1-FOR-57	TAATACGACTCACTATAGCCGGGCCTTTCCTGACAACCCGATGCTGACACCTCCCCAAGAAACCA
Tktl1-FOR-59	TAATACGACTCACTATAGCCGTTCCGTGTCTGAGTCGAGATCCTGCCCAGCACAGCTCTCCCCCA
Tktl1-FOR-61	TAATACGACTCACTATAGCCGGAGTTCTGCTCTTAGGACCCAAAATGAAAGAGAATACTGTCCCA
Tktl1-FOR-63	TAATACGACTCACTATAGCCTGGAAAGAGACTACCAGATCCTTAGTAAGCCTCCATCAATTCCCA
Tktl1-FOR-65	TAATACGACTCACTATAGCCAAATGGAACAACTAAGAAGGATGAGGGATAGGAGCTGGGCTGCCA
Tktl1-FOR-67	TAATACGACTCACTATAGCCGTATCTTCCCCCGATTGGGAAGAGGTCTTGTGCCTGACATCTCCA
Tktl1-FOR-69	TAATACGACTCACTATAGCCCAGAGGTTGGAGGGCCGTGACAAAGGAACTTCGGAGTCCCTACCA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Tktl1-FOR-71	TAATACGACTCACTATAGCCCGGTGACGCAGAACGATTTTCCCATTTGCACCATCCCCTCTGCCA
Tktl1-FOR-74	TAATACGACTCACTATAGCCAAGTCAAACCCCGATTAATGGACAGAGGATTGCTTCCTCCATCCA
Tktl1-FOR-76	TAATACGACTCACTATAGCCACCGATCAGCGAAGATAGCAGTGTCTGCCCGATATTGGTTTGCCA
Tktl1-FOR-78	TAATACGACTCACTATAGCCGGGAGTAGTACGAAAACTGCTTTTTGGGATGGGATAGGAAACCCA
Tktl1-FOR-81	TAATACGACTCACTATAGCCGGCGACTTTGGCGGTGACGGAAGCCTGAAACCGCTGCTGCCACCA
Tktl1-FOR-84	TAATACGACTCACTATAGCCAAGGTACAGCTTGTGTCATCTTGTGTTTCCTAAATGCATTTCCCA
Tktl1-FOR-86	TAATACGACTCACTATAGCCTCGAATCAAGATTCATTCATTTATGCATTCCAGGACTCTAGCCCA
Tktl1-FOR-88	TAATACGACTCACTATAGCCTAGCGTTCCTCGTGATAGACAGTCCGGCCCAGGAGCCTGATTCCA
Tktl1-FOR-91	TAATACGACTCACTATAGCCAACCCCCGACCCCTCTCACATTTTTTGTTTGACCTCCCATTTCCA
Tktl1-FOR-93	TAATACGACTCACTATAGCCTCGTTCATAACGGCACTCTCCAAAGAGGCTACAATACAAGAGCCA
Tktl1-FOR-95	TAATACGACTCACTATAGCCCGACCTCTCTACGCTTAGTGTTTAAGAGCACCAGTTGTTCTTCCA
Tktl1-FOR-99	TAATACGACTCACTATAGCCAAAAGTAGCAATATAATGATTCCTAATGATATTCTGCCATACCCA
Tktl1-FOR-101	TAATACGACTCACTATAGCCCCGGTCCGTTGACGACCGAGCTACCTCTGCTTTTCTTGACCACCA
Tktl1-FOR-104	TAATACGACTCACTATAGCCAACAAGATTGGACTAGGTCCTGTTGTAGCTCTCCTATACTCACCA
Tktl1-FOR-108	TAATACGACTCACTATAGCCCGTAGACGTATCCCTGTAATGAAGTAGAGGGGACAGGCCATACCA
Tktl1-FOR-111	TAATACGACTCACTATAGCCATAACGATCACGAGGAAAGAGACAGCAGGCAGCCAGAAGCCTCCA
Tktl1-FOR-113	TAATACGACTCACTATAGCCAGACCACTCCTCCTTACACTAGTAGCACTGAGGCTAGTGTGGCCA
Tktl1-FOR-115	TAATACGACTCACTATAGCCAGAGTAGTGCCTGTAGTACTGAACCTGTAGTGGCAGGCTTTTCCA
Tktl1-REV-3	TGGCCACCCAGTTTCGTTTTCTTTCCGAAAACGAGGAGACTCGGATCCCTTTAGTGAGGGTTAATA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Tktl1-REV-8	TGGAAAAGCGGGTAGGAGCTGACTGGGAGAGAGTCGAAGTCGGAATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-10	TGGAACACACATTCTGGAAGGGAACTCACTACAGAGTGTTCAAGATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-14	TGGACCTCTTGGTGAATGGATATTTTGTTCAATTCATTTTTATGATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-16	TGGGAGGGCTGTACTGGGCCTCCCTCTTCCGTCTTATTCCCTCTCTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-19	TGGCTTTACTTTGTACCTTGGCTTGGGGACTGTATCATTGATGGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-21	TGGCAATTGGGCAGAAGCCTCAGGATCGTCTCCTCTTCACTCCGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-24	TGGCCAAATTTGGAGGTGGTGTCAATGAAGAAGTCCAGTTAGAAGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-31	TGGAGAGAATGTGCCCTCTCTACAGGACTGTTTCGGTACGGTTTCTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-35	TGGATAAATGTTTTGAATACTCACAGGAAGGAAATACTTACTTGGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-37	TGGGAGATGGGCAAAACTGCTGAGGAACCTCTGCTTAAATAATATTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-39	TGGAGACTGCTGCACAGACAGCTCGAGGTTACGGTGGATGTCTTTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-44	TGGTGGCTTTGGATTGGAGGGATTTTGTCCGAACTAGGATCCCGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-46	TGGCCTATGCTGAAATTCTGCCAGGTAAGTGAAGTTTCCTGAAACTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-48	TGGCAATGAGCATCCCCACACCTCAGGATAAGGGTTCTTCGGGGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-50	TGGAGTGGTCAACAAACCTGCCACGAAGTGACAGTTGTGGTTCCTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-52	TGGGCCTGGCCAGAGTGCCGCATATTCAATCGTATCCGGTACTATTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-54	TGGGGGGTTGAGGGTCCCAGTTCGATAAAGGGACAATATGGTCATTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-56	TGGTGATGGCACACACACCATTACCTATATTCGAGAGACAGGACCTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-58	TGGAGGGTGTCCATACAGTACATGTCACCAAGCGGCCACAAGGGTTCCCTTTAGTGAGGGTTAATA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Tktl1-REV-60	TGGGTAGTTGGTGAAGGGCCCCAATGGAGTGAGAGTGTCGGACGGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-62	TGGAGCAAGATCTGACTTGTGCGGGGTTGGGTCCGAACGAGTCTCTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-64	TGGCAGACAAGGCCGGCACCACCTTCGCCACAACTTGCGGTCCCCTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-66	TGGGTGCACTGTAGTCTTTGATTCTACCACGTACGGTGGTTAGGGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-68	TGGTGCCGCTCCCTGGTTACATCACGGGTCGAACAAGGTCGTGTATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-70	TGGGAAAAGGGGACTTGGTCGGGAATTAATCGACGTGTTTTTTAGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-72	TGGTTAAGCTTCTCCAAAATGAAGTCCATCCCTTAGACCATCTGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-75	TGGTTATATCACGGGCCACCTGTAGAAACATCGTTGTTCGGAAGATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-77	TGGTCTTTAGGCTGAGTATCATGTGCAGACCTATCCACTTATACTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-80	TGGGAAGCCAATTGGTTAACATGTACCAGAGCTTTTACCGGTGTGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-83	TGGCACCATCATACCTCATCCTTCTCCAGTCACCGTAGTGACTCGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-85	TGGAACACCATTCTATGTCTTCATTGGCCACTTCTTACACTGGAGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-87	TGGATCAAGACTTCCTTTTGGGGCCAAATAATATCGACAACGTACTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-89	TGGGAAAATGGGACTGGGGATGGAAAGAGCCCCGTTCCTCCTCTTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-92	TGGTGAGTGAGGCCACAGTACCACCCAGTGCGTACCTCTTTGATGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-94	TGGGGCATATATATTTACATATGGAAATTAAAGTGGAATTACATTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-97	TGGTCCTAAATGCTAAGATTTCTTCAGGCATCACCTGTGGTCCGGTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-100	TGGGGAGCTCAAGTGATGGCCAGACGGTTGTAGTCGTGACGAGACTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-103	TGGTGGTATCTTTGCATCTTGAAATGAGAGTGGCTACCGGATCACTCCCTTTAGTGAGGGTTAATA

Continued on next page

Table A.1.4 – *Continued from previous page*

Primer	Sequence
Tktl1-REV-107	TGGGGAGAGACAGTAACTTTGATCTCTCCTGAGGACATATATAAATCCCTTTAGTGAGGGTTAATA
Tktl1-REV-109	TGGGAGCCCTACCCTTTTCTGGAGCTCACCTATCCTTCCCCCAGTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-112	TGGTCAACCTCATCCCACCTATGGGGTCGGAGGAATGGTTTCCTTTCCCTTTAGTGAGGGTTAATA
Tktl1-REV-114	TGGGTCCACCACTGCCACTTTTGGTCGTCGACAGGAGATACACCTTCCCTTTAGTGAGGGTTAATA

Table A.1.5: 4C Primers

Primer	Enzyme	Sequence
Xlr3b-F1	HindIII-DpnII	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTCTGTTTGCGCAAGCTT
Xlr3b-R1	HindIII-DpnII	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTCTGCAACTTCTGCAAT
F8a-F2	HindIII-DpnII	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAAGGCTTACCAAAGCTT
F8a-R2	HindIII-DpnII	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAACATTTGGCTCACATCTC
Xlr4bc-F3	HindIII-DpnII	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCAGCAGGTGAAAAGCTT
Xlr4bc-R3	HindIII-DpnII	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCACAACCAGGACTGTATATAAG
Xlr3b-F4	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTGACTGCACACCTGGATC
Xlr3b-R4	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCACTTGGTCGGGCCTT
F8a-F5	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACCTTTGTCGGAAATGATC
F8a-R5	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGCCTGCCTCTTCTATTTA
Xlr4c-F6	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGGTGGAATTAAGAAGATC

Continued on next page

Table A.1.5 – *Continued from previous page*

Primer	Enzyme	Sequence
Xlr4c-R6	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTGGGGCTTACTCTACTCC
Syn1-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGAGGAGACTGGGCAGATC
Syn1-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTAGGATGTCTCTGGGTCTGG
Cul4b-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTGCTTAGGAGTGGAGATC
Cul4b-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCCCTGAGTTTTTGATTTGAAG
Fmr1-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTAAATAGCAGTCTTTGGATC
Fmr1-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCATCATAGCATCCCTTCATT
MeCP2-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTAACAAAAAGGTCAAAAGGATC
MeCP2-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAATACCCCCTGCTTTTAGG
Nlgn3-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCAGGCAGTTTAGGGGATC
Nlgn3-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACTCTCCCAGGAGAAAC
Ap1s2-F	DpnII-Csp6I	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGATGGTCTATGTGTCCAGGATC
Ap1s2-R	DpnII-Csp6I	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTTATTTCTTCCCCATTTCTGG

A.2 BACs

Table A.2.6: BAC Clones

BAC	Coordinates (mm9)	Length (bp)
RP23-170B5	chrX:70223278-70271621	48344
RP23-11C16	chrX:70,269,631-70,475,959	206329
RP23-308D2	chrX:70,476,166-70,690,675	214510
RP23-171N20	chrX:71350711-71464953	114243
RP23-436K3	chrX:71,462,970-71,666,091	203122

A.3 Antibodies

Table A.3.7: ChIP Antibodies

Antibody	Supplier	Catalogue #
ATRX (D-5) Monoclonal	Santa Cruz	sc-55584
H3K27me3 Polyclonal (Classic)	Diagenode	C15410069
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5)	Abcam	ab5131
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2)	Abcam	ab5095
Anti-CTCF	Abcam	ab70303

Appendix B

B.1 Sequencing Metrics

Table B.1.8: 5C-Seq Library Metrics

Library	Total Reads	Filtered Reads	Interactions
39,X^M	5902600	4869139	7223
39,X^P	5960433	4985697	6589
BAC	4468148	4156345	7655

Table B.1.9: 4C-Seq Library Metrics

Bait	Library	Replicate	Total Reads	Filtered Reads	Reads in Cis
Xlr3b (HindIII)	39,X^M	Rep1-136	1511491	864245	231906 (26.83%)
		Rep2-144	1259259	688926	180089 (26.14%)
		Rep3-149	1381160	701660	217712 (31.03%)
	39,X^P	Rep1-745	1370660	827640	177666 (21.47%)

Continued on next page

Table B.1.9 – *Continued from previous page*

Bait	Library	Replicate	Total Reads	Filtered Reads	Reads in Cis
		Rep2-820	12859969	7849081	1621828 (20.66%)
		Rep3-826	837265	504887	121551 (24.07%)
F8a (HindIII)	39,X^M	Rep1-136	2480900	1010925	175498 (17.36%)
		Rep2-144	3438861	1364425	174304 (12.77%)
		Rep3-149	3899141	1527960	272576 (17.84%)
	39,X^P	Rep1-745	2472330	1038446	132786 (12.79%)
		Rep2-820	3002884	1477714	184006 (12.45%)
		Rep3-826	2797150	911554	170337 (18.69%)
Xlr4bc (HindIII)	39,X^M	Rep1-136	2585288	1650824	822082 (49.8%)
		Rep2-144	2494637	1562011	801748 (51.33%)
		Rep3-149	2213384	1338435	762391 (56.96%)
	39,X^P	Rep1-745	2808116	1833240	788797 (43.03%)
		Rep2-820	2587810	1750071	654069 (37.37%)
		Rep3-826	3274956	2135574	976822 (45.74%)
Xlr3b (DpnII)	39,X^M	Rep1-128	4362105	1925579	998903 (51.88%)
		Rep2-133	5787954	2491944	1274993 (51.16%)
		Rep3-161	5964820	2630753	1372002 (52.15%)
	39,X^P	Rep1-261	4438329	2084245	971787 (46.63%)
		Rep2-268	4690676	2083538	1090752 (52.35%)

Continued on next page

Table B.1.9 – *Continued from previous page*

Bait	Library	Replicate	Total Reads	Filtered Reads	Reads in Cis
		Rep3-772	2939814	1491241	783585 (52.55%)
F8a (DpnII)	39,X^M	Rep1-128	3960812	2979126	1555187 (52.2%)
		Rep2-133	3272147	2532640	1318950 (52.08%)
		Rep3-161	2587695	2043218	1051996 (51.49%)
	39,X^P	Rep1-261	3623627	2666497	1204921 (45.19%)
		Rep2-268	3447529	2728263	1363003 (49.96%)
		Rep3-772	3125973	2458364	1232937 (50.15%)
Xlr4c (DpnII)	39,X^M	Rep1-128	1302169	808291	448225 (55.45%)
		Rep2-133	1462757	910775	499513 (54.84%)
		Rep3-161	1249014	819599	454548 (55.46%)
	39,X^P	Rep1-261	1586892	953193	449856 (47.19%)
		Rep2-268	1437939	916676	475215 (51.84%)
		Rep3-772	1431044	879313	463328 (52.69%)
Syn1 (DpnII)	39,X^M	Rep1-247	5079845	2416810	1537494 (63.62%)
		Rep2-256	7547572	3285533	1607406 (48.92%)
		Rep3-280	10289737	3091566	1920152 (62.11%)
	39,X^P	Rep1-907	8947149	4438755	1533690 (34.55%)
		Rep2-913	6710115	1874922	1138436 (60.72%)
		Rep3-941	8212261	3855391	1733289 (44.96%)

Continued on next page

Table B.1.9 – *Continued from previous page*

Bait	Library	Replicate	Total Reads	Filtered Reads	Reads in Cis
Cul4b (DpnII)	39,X^M	Rep1-247	3233893	1719372	1231435 (71.62%)
		Rep2-256	4411943	2155983	1196279 (55.49%)
		Rep3-280	8669928	4178712	2516968 (60.23%)
	39,X^P	Rep1-907	7370880	5147344	1855688 (36.05%)
		Rep2-913	3796261	1547998	818232 (52.86%)
		Rep3-941	9318181	6790180	2783470 (40.99%)
Fmr1 (DpnII)	39,X^M	Rep1-256	8295972	2184131	1416438 (64.85%)
		Rep2-260	6903990	1374235	882952 (64.25%)
		Rep3-280	8843517	2030107	1273029 (62.71%)
	39,X^P	Rep1-907	8508069	2712871	1083852 (39.95%)
		Rep2-913	5586939	741034	423388 (57.13%)
		Rep3-941	8083571	2321154	1095607 (47.2%)
MeCP2 (DpnII)	39,X^M	Rep1-260	2426830	1623899	927145 (57.09%)
		Rep2-275	1716022	609642	565535 (92.77%)
		Rep3-280	5590431	3851383	2365043 (61.41%)
	39,X^P	Rep1-907	6113024	4787039	1707959 (35.68%)
		Rep2-913	5032572	2900292	1508100 (52.0%)
		Rep3-941	6465281	4892106	2149956 (43.95%)
Nlgn3 (DpnII)	39,X^M	Rep1-260	6006475	4167816	2140181 (51.35%)

Continued on next page

Table B.1.9 – *Continued from previous page*

Bait	Library	Replicate	Total Reads	Filtered Reads	Reads in Cis
		Rep2-275	6930065	4179751	2331832 (55.79%)
		Rep3-280	7726404	6020690	3274795 (54.39%)
	39,X^P	Rep1-907	7277074	5835211	2050859 (35.15%)
		Rep2-913	2266323	1836647	886835 (48.29%)
		Rep3-941	6324381	4960396	2085531 (42.04%)
Apls2 (DpnII)	39,X^M	Rep1-260	7419412	2700726	1664857 (61.64%)
		Rep2-275	8453155	2163297	1320581 (61.04%)
		Rep3-280	9105942	2834692	1615586 (56.99%)
	39,X^P	Rep1-907	7190506	3470162	1194227 (34.41%)
		Rep2-913	6443400	1670731	922373 (55.21%)
		Rep3-941	8223202	2817058	1390756 (49.37%)

Table B.1.10: ChIP-Seq Library Metrics

ChIP	Library	Replicate	Total Reads	Filtered Reads
Atrx	39,X^M	Rep1	99525482	92768826
		Rep2	85011980	78980053
	39,X^P	Rep1	74323130	67019945
		Rep2	89216790	79604790
H3K27me3	39,X^M	Rep1	114504982	110089268
		Rep2	94280854	87651968
	39,X^P	Rep1	113446706	109037547
		Rep2	55497628	49170107
Input	39,X^M		95230152	89530078
	39,X^P		231149838	126975600

Bibliography

[1] Kanherkar, Riya R and Bhatia-Dey, Naina and Csoka, Antonei B, *Epigenetics across the human lifespan*, Front Cell Dev Biol, vol. 2, p. 49, 2014.

[2] Phillips, Theresa, *The role of methylation in gene expression*, Nature Education, vol. 1(1), p. 116, 2008.

[3] Robertson, Keith D (2005), *Epigenetic Mechanisms of Gene Regulation*, In: DNA Methylation and Cancer Therapy. Medical Intelligence Unit. Springer, Boston, MA.

[4] Gendrel, Anne-Valerie and Heard, Edith, *Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation*, Annu Rev Cell Dev Biol, vol. 30, p. 561-80, 2014.

[5] Wood, Andrew J and Oakey, Rebecca J, *Genomic imprinting in mammals: emerging themes and established theories*, PLoS Genet, vol. 2(11), e147, Nov 2006.

[6] Inoue, Azusa and Jiang, Lan and Lu, Falong and Suzuki, Tsukasa and Zhang, Yi, *Maternal H3K27me3 controls DNA methylation-independent imprinting*, Nature, vol. 547(7664), p. 419-24, July 2017.

[7] Inoue, Azusa and Jiang, Lan and Lu, Falong and Zhang, Yi, *Genomic imprinting of Xist by maternal H3K27me3*, Genes Dev, vol. 31(19), p. 1927-32, Oct 2017.

[8] Skuse, D H and James, R S and Bishop, D V and Coppin, B and Dalton, P and Aamodt-Leeper, G and Bacarese-Hamilton, M and Creswell, C and McGurk, R and Jacobs, P A, *Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function*, Nature, vol. 387(6634), p. 705-8, Jun 1997.

[9] Fombonne, Eric, *Epidemiology of pervasive developmental disorders*, Pediatr Res, vol. 65(6), p. 591-8, Jun 2009.

[10] Bajrami, Emirjeta and Spiroski, Mirko, *Genomic Imprinting*, Open Access Maced J Med Sci, vol. 4(1), p. 181-4, Mar 2016.

[11] Joyce, J A and Schofield, P N, *Genomic imprinting and cancer*, Mol Pathol, vol. 51(4), p. 185-90, Aug 1998.

[12] Kochanski, A and Merritt, T A and Gadzinowski, J and Jopek, A, *The impact of assisted reproductive technologies on the genome and epigenome of the newborn*, J Neonatal Perinatal Med, vol. 6(2), p. 101-8, 2013.

[13] Takagi, N and Sasaki, M, *Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse*, Nature, vol. 256(5519), p. 640-2, Aug 1975.

[14] Surani, M A and Barton, S C and Norris, M L, *Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis*, Nature, vol. 308(5959), p. 548-50, 1984.

[15] Haig, D, *The Kinship Theory of Genomic Imprinting*, Annu Rev Ecol Evol Syst, vol. 31(1), p. 9-32, Nov 2000.

[16] Latham, Keith E, *X chromosome imprinting and inactivation in preimplantation mammalian embryos*, Trends Genet, vol. 21(2), p. 120-7, Feb 2005.

[17] Reik, Wolf and Lewis, Annabelle, *Co-evolution of X-chromosome inactivation and imprinting in mammals*, Nat Rev Genet, vol. 6(5), p. 403-10, May 2005.

[18] Chang, Samuel C and Tucker, Tracy and Thorogood, Nancy P and Brown, Carolyn J, *Mechanisms of X-chromosome inactivation*, Front Biosci, vol. 11, p. 852-66, Jan 2006.

[19] Ohlsson, R and Paldi, A and Graves, J A, *Did genomic imprinting and X chromosome inactivation arise from stochastic expression?*, Trends Genet, vol. 17(3), p. 136-41, Mar 2001.

[20] Ranke, M B and Saenger, P, *Turner's syndrome*, Lancet, vol. 358(9278), p. 309-14, Jul 2001.

[21] Lepage, Jean-Francois and Hong, David S and Mazaika, Paul K and Raman, Mira and Sheau, Kristen and Marzelli, Matthew J and Hallmayer, Joachim and Reiss, Allan L, *Genomic imprinting effects of the X chromosome on brain morphology*, J Neurosci, vol. 33(19), p. 8567-74, May 2013.

[22] Lehrke, R, *Theory of X-linkage of major intellectual traits*, Am J Ment Defic, vol. 76(6), p. 611-9, May 1972.

[23] Ross, Judith and Roeltgen, David and Zinn, Andrew, *Cognition and the sex chromosomes: studies in Turner syndrome*, Horm Res, vol. 65(1), p. 47-56, 2006.

[24] Gockley, Jake and Willsey, A Jeremy and Dong, Shan and Dougherty, Joseph D and Constantino, John N and Sanders, Stephan J, *The female protective effect in autism spectrum disorder is not mediated by a single genetic locus*, Mol Autism, vol. 6, p. 25, 2015.

[25] Donnelly, S L and Wolpert, C M and Menold, M M and Bass, M P and Gilbert, J R and Cuccaro, M L and Delong, G R and Pericak-Vance, M A, *Female with autistic disorder and monosomy X (Turner syndrome): parent-of-origin effect of the X chromosome*, Am J Med Genet, vol. 2000(96), p. 312-6, Jun 2000.

[26] Bartolomei, Marisa S and Ferguson-Smith, Anne C, *Mammalian genomic imprinting*, Cold Spring Harb Perspect Biol, vol. 3(7), Jul 2011.

[27] Veltman, Marijcke W M and Craig, Ellen E and Bolton, Patrick F, *Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review*, Psychiatr Genet, vol. 15(4), p. 243-54, Dec 2005.

[28] Cassidy, S B and Dykens, E and Williams, C A, *Prader-Willi and Angelman syndromes: sister imprinted disorders*, Am J Med Genet, vol. 97(2), p. 136-46, 2000.

[29] Thomson, J A and Solter, D *The developmental fate of androgenetic, parthenogenetic, and gynogenetic cells in chimeric gastrulating mouse embryos*, Genes Dev, vol. 2(10), p. 1344-51, Oct 1988.

[30] Wilkinson, Lawrence S and Davies, William and Isles, Anthony R, *Genomic imprinting effects on brain development and function*, Nat Rev Reurosci, vol. 8(11), p. 832-43, Nov 2007.

[31] Gejman, Pablo V and Sanders, Alan R and Duan, Jubao, *The role of genetics in the etiology of schizophrenia*, Psychiatr Clin North Am, vol. 33(1), p. 35-66, Mar 2010.

[32] Raefski, Adam S and O'Neill, Michael J, *Identification of a cluster of X-linked imprinted genes in mice*, Nat Genet, vol. 37(6), p. 620-4, Jun 2005.

[33] Davies, William and Isles, Anthony and Smith, Rachel and Karunadasa, Delicia and Burrmann, Doreen and Humby, Trevor and Ojarikre, Obah and Biggin, Carol and Skuse, David and Burgoyne, Paul and Wilkinson, Lawrence, *Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice*, Nat Genet, vol. 37(6), p. 625-9, Jun 2005.

[34] Kobayashi, Shin and Isotani, Ayako and Mise, Nathan and Yamamoto, Masamichi and Fujihara, Yoshitaka and Kaseda, Kazuhiro and Nakanishi, Tomoko and Ikawa, Masahito and Hamada, Hiroshi and Abe, Kuniya and Okabe, Masaru, *Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation stages*, Curr Biol, vol. 16(2), p. 166-72, Jan 2006.

[35] Kobayashi, Shin and Fujihara, Yoshitaka and Mise, Nathan and Kaseda, Kazuhiro and Abe, Kuniya and Ishino, Fumitoshi and Okabe, Masaru, *The*

X-linked imprinted gene family Fthl17 shows predominantly female expression following the two-cell stage in mouse embryos, Nucleic Acids Res, vol. 38(1), p. 3672-81, Jun 2010.

[36] Dobson, M J and Pearlman, R E and Karaiskakis, A and Spyropoulos, B and Moens, P B, *Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction*, J Cell Sci, vol. 107(10), p. 2749-60, Oct 1994.

[37] Kelsey, Gavin and Bartolomei, Marisa S, *Imprinted genes ? and the number is?* PLoS Genetics, vol. 8(3), e1002601, 2012.

[38] Kurukuti, Sreenivasulu and Tiwari, Vijay Kumar and Tavoosidana, Gholamreza and Pugacheva, Elena and Murrell, Adele and Zhao, Zhihu and Lobanenkov, Victor and Reik, Wolf and Ohlsson, Rolf, *CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2*, Proc Natl Acad Sci U S A, vol. 103(28), p. 10684-9, Jul 2006.

[39] Jackson-Grusby, L and Beard, C and Possemato, R and Tudor, M and Fambrough, D and Csankovszki, G and Dausman, J and Lee, P and Wilson, C and Lander, E and Jaenisch, R, *Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation*, Nat Genet, vol. 27(1), p. 31-9, Jan 2001.

[40] Schulz, Reiner and Woodfine, Kathryn and Menheniott, Trevelyan R and Bourc'his, Deborah and Bestor, Timothy and Oakey, Rebecca J, *WAMIDEX: a web atlas of murine genomic imprinting and differential expression*, Epigenetics, vol. 3(2), p. 89-96, 2008.

[41] Prendergast, James G D and Tong, Pin and Hay, David C and Farrington, Susan M and Semple, Colin A M, *A genome-wide screen in human embryonic stem cells reveals novel sites of allele-specific histone modification associated with known disease loci*, Epigenetics Chromatin, vol. 5(1), p. 6, May 2012.

[42] Zupkovitz, Gordin and Tischler, Julia and Posch, Markus and Sadzak, Iwona and Ramsauer, Katrin and Egger, Gerda and Grausenburger, Reinhard and Schweifer, Norbert and Chiocca, Susanna and Decker, Thomas and Seiser, Christian, *Negative and positive regulation of gene expression by mouse histone deacetylase 1*, Mol Cell Biol, vol. 26(21), p. 7913-28, Nov 2006.

[43] Fan, Yuhong and Nikitina, Tatiana and Zhao, Jie and Fleury, Tomara J and Bhattacharyya, Riddhi and Bouhassira, Eric E and Stein, Arnold and Woodcock, Christopher L and Skoultchi, Arthur I, *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation*, Cell, vol. 123(7), p. 1199-212, Dec 2005.

[44] Zhao, Zhihu and Tavoosidana, Gholamreza and Sjolinder, Mikael and Gondor, Anita and Mariano, Piero and Wang, Sha and Kanduri, Chandrasekhar and Lezcano, Magda and Sandhu, Kuljeet Singh and Singh, Umashankar and Pant, Vinod and Tiwari, Vijay and Kurukuti, Sreenivasulu and Ohlsson, Rolf, *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions*, Nat Genet, vol. 38(11), p. 1341-7, Nov 2006.

[45] Murrell, Adele and Heeson, Sarah and Reik, Wolf, *Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops*, Nat Genet, vol. 36(8), p. 889-93, Aug 2004.

[46] Braem, Caroline and Recolin, Benedicte and Rancourt, Rebecca C and Angiolini, Christopher and Barthes, Pauline and Branchu, Priscillia and Court, Franck and Cathala, Guy and Ferguson-Smith, Anne C and Forne, Thierry, *Genomic matrix attachment region and chromosome conformation capture quantitative real time PCR assays identify novel putative regulatory elements at the imprinted Dlk1/Gtl2 locus*, J Biol Chem, vol. 283(27), p. 18612-20, Jul 2008.

[47] Lonfat, Nicolas and Montavon, Thomas and Jebb, David and Tschopp, Patrick and Nguyen Huynh, Thi Hanh and Zakany, Jozsef and Duboule, Denis, *Transgene- and locus-dependent imprinting reveals allele-specific chromosome conformations*, Proc Natl Acad Sci U S A, vol. 110(29), p. 11946-51, Jul 2013.

[48] Horike, Shin-ichi and Cai, Shutao and Miyano, Masaru and Cheng, Jan-Fang and Kohwi-Shigematsu, Terumi, *Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome*, Nat Genet, vol. 37(1), p. 31-40, Jan 2005.

[49] Cremer, Thomas and Cremer, Marion, *Chromosome territories*, Cold Spring Harb Perspect Biol, vol. 2(3), p. a003889, Mar 2010.

[50] Cremer, T and Cremer, C, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells*, Nat Rev Genet, vol. 2(4), p. 292-301, Apr 2001.

[51] de Wit, Elzo and de Laat, Wouter, *A decade of 3C technologies: insights into nuclear organization*, Genes Dev, vol. 26(1), p. 11-24, Jan 2012.

[52] Lieberman-Aiden, Erez and van Berkum, Nynke L and Williams, Louise and Imakaev, Maxim and Ragoczy, Tobias and Telling, Agnes and Amit, Ido and Lajoie, Bryan R and Sabo, Peter J and Dorschner, Michael O and Sandstrom, Richard and Bernstein, Bradley and Bender, M A and Groudine, Mark and Gnirke, Andreas and Stamatoyannopoulos, John and Mirny, Leonid A and Lander, Eric S and Dekker, Job, *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*, Science, vol. 326(5950), p. 289-93, Oct 2009.

[53] Simonis, Marieke and Klous, Petra and Splinter, Erik and Moshkin, Yuri and Willemsen, Rob and de Wit, Elzo and van Steensel, Bas and de Laat, Wouter, *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)*, Nat Genet, vol. 38(11), p. 1348-54, Nov 2006.

[54] Court, Franck and Baniol, Marion and Hagege, Helene and Petit, Julie Sandrine and Lelay-Taha, Marie-Nolle and Carbonell, Franoise and Weber, Michael and Cathala, Guy and Forne, Thierry, *Long-range chromatin interactions at the mouse Igf2/H19 locus reveal a novel paternally expressed long non-coding RNA*, Nucleic Acids Res, vol. 39(14), p. 5893-906, Aug 2011.

[55] Tolhuis, Bas and Palstra, Robert Jan and Splinter, Erik and Grosveld, Frank and de Laat, Wouter, *Looping and interaction between hypersensitive sites in the active beta-globin locus*, Mol Cell, vol. 10(6), p. 1453-65, Dec 2002.

[56] Phillips-Cremins, Jennifer E and Sauria, Michael E G and Sanyal, Amartya and Gerasimova, Tatiana I and Lajoie, Bryan R and Bell, Joshua S K and Ong, Chin-Tong and Hookway, Tracy A and Guo, Changying and Sun, Yuhua and Bland, Michael J and Wagstaff, William and Dalton, Stephen and McDevitt, Todd C and Sen, Ranjan and Dekker, Job and Taylor, James and Corces, Victor G, *Architectural protein subclasses shape 3D organization of genomes during lineage commitment*, Cell, vol. 153(6), p. 1281-95, Jun 2013.

[57] de Wit, Elzo and Vos, Erica S M and Holwerda, Sjoerd J B and Valdes-Quezada, Christian and Verstegen, Marjon J A M and Teunissen, Hans and Splinter, Erik and Wijchers, Patrick J and Krijger, Peter H L and de Laat, Wouter, *CTCF Binding Polarity Determines Chromatin Looping*, Mol Cell, vol. 60(4), p. 676-84, Nov 2015.

[58] Guo, Ya and Xu, Quan and Canzio, Daniele and Shou, Jia and Li, Jinhuan and Gorkin, David U and Jung, Inkyung and Wu, Haiyang and Zhai, Yanan and Tang, Yuanxiao and Lu, Yichao and Wu, Yonghu and Jia, Zhilian and Li, Wei and Zhang, Michael Q and Ren, Bing and Krainer, Adrian R and Maniatis, Tom and Wu, Qiang, *CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function*, Cell, vol. 162(4), p. 900-10, Aug 2015.

[59] Nora, Elphege P and Goloborodko, Anton and Valton, Anne-Laure and Gibcus, Johan H and Uebersohn, Alec and Abdennur, Nezar and Dekker, Job and Mirny, Leonid A and Bruneau, Benoit G, *Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization*, Cell, vol. 169(5), p. 930-944, May 2017.

[60] Vu, Thanh H and Nguyen, An H and Hoffman, Andrew R, *Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells*, Hum Mol Genet, vol. 19(5), p. 901-19, Mar 2010.

[61] van de Werken, Harmen J G and de Vree, Paula J P and Splinter, Erik and Holwerda, Sjoerd J B and Klous, Petra and de Wit, Elzo and de Laat, Wouter, *4C technology: protocols and data analysis*, Methods Enzymol, vol. 513, p. 89-112, 2012.

[62] Noordermeer D, *Ensembles of mono- and bi-allelic CTCF sites structure differential sub-TAD organization at paternally imprinted gene loci*, Poster presented at: CSHL Nuclear Organization & Function; 2018 May 1-5; Cold Spring Harbor, NY.

[63] Ryba, Tyrone and Hiratani, Ichiro and Lu, Junjie and Itoh, Mari and Kulik, Michael and Zhang, Jinfeng and Schulz, Thomas C and Robins, Allan J and Dalton, Stephen and Gilbert, David M, *Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types*, Genome Res, vol. 20(6), p. 761-70, Jun 2010.

[64] Cai, Mingyang and Gao, Fan and Lu, Wange and Wang, Kai, *w4CSeq: software and web application to analyze 4C-seq data*, Bioinformatics, vol. 32(21), p. 3333-5, Nov 2016.

[65] Bannister, Andrew J and Kouzarides, Tony, *Regulation of chromatin by histone modifications*, Cell Res, vol. 21(3), p. 381-95, Mar 2011.

[66] Kagey, Michael H and Newman, Jamie J and Bilodeau, Steve and Zhan, Ye and Orlando, David A and van Berkum, Nynke L and Ebmeier, Christopher C and Goossens, Jesse and Rahl, Peter B and Levine, Stuart S and Taatjes, Dylan J and Dekker, Job and Young, Richard A, *Mediator and cohesin connect gene expression and chromatin architecture*, Nature, vol. 467(7314), p. 430-5, Sep 2010.

[67] Kernohan, Kristin D and Jiang, Yan and Tremblay, Deanna C and Bonvissuto, Anne C and Eubanks, James H and Mann, Mellissa R W and Brub, Nathalie G, *ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain*, Dev Cell, vol. 18(2), p. 191-202, Feb 2010.

[68] Levy, Michael A and Kernohan, Kristin D and Jiang, Yan and Berube, Nathalie G, *ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions*, Hum Mol Genet, vol. 24(7), p. 1824-35, Apr 2015.

[69] Bailey, J A and Carrel, L and Chakravarti, A and Eichler, E E, *Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis*, Proc Natl Acad Sci U S A, vol. 97(12), p. 6634-39, Jun 2000.

[70] Ciabrelli, Filippo and Cavalli, Giacomo, *Chromatin-driven behavior of topologically associating domains*, J Mol Biol, vol. 427(3), p. 608-25, Feb 2015.

[71] Rao, Suhas S P and Huntley, Miriam H and Durand, Neva C and Stamenova, Elena K and Bochkov, Ivan D and Robinson, James T and Sanborn, Adrian

L and Machol, Ido and Omer, Arina D and Lander, Eric S and Aiden, Erez Lieberman, *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping*, Cell, vol. 159(7), p.1665-80, Dec 2014.

[72] McLean, Cory Y and Bristor, Dave and Hiller, Michael and Clarke, Shoa L and Schaar, Bruce T and Lowe, Craig B and Wenger, Aaron M and Bejerano, Gill, *GREAT improves functional interpretation of cis-regulatory regions*, Nat Biotechnol, vol. 28(5), p. 495-501, May 2010.

[73] Kohler, Sebastian and Carmody, Leigh and Vasilevsky, Nicole and Jacobsen, Julius O B, et al, *Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources*, Nucleic Acids Res. (2018) doi:10.1093/nar/gky1105.

[74] McElreavey, K and Barbaux, S and Ion, A and Fellous, M, *The genetic basis of murine and human sex determination: a review*, Heredity (Edinb), vol. 75, p. 599-611, Dec 1995.

[75] Graves, Jennifer A Marshall, *Sex chromosome specialization and degeneration in mammals*, Cell, vol. 124(5), p. 901-14, Mar 2006.

[76] Muller, H J, *A gene for the fourth chromosome of Drosophila*, J Exp Zool, vol. 17, p. 325-36, Oct 1914.

[77] Quintana-Murci, Lluis and Fellous, Marc, *The Human Y Chromosome: The Biological Role of a "Functional Wasteland"* J Biomed Biotechnol, vol. 1(1), p. 18-24, 2001.

[78] Giannelli, F and Green, P M, *The X chromosome and the rate of deleterious mutations in humans*, Am J Hum Genet, vol. 67(2), p. 515-17, Aug 2000.

[79] Saifi, G M and Chandra, H S, *An apparent excess of sex- and reproduction-related genes on the human X chromosome*, Proc Biol Sci, vol. 266(1415), p. 203-9, Jan 1999.

[80] Johnson, Norman A and Lachance, Joseph *The genetics of sex chromosomes: evolution and implications for hybrid incompatibility*, Ann N Y Acad Sci, vol. 1256, p. E1-22, May 2012.

[81] Rice, William R, *Sex-chromosomes and the evolution of sexual dimorphism*, Evolution, vol. 38(4), p. 735-42, Jul 1984.

[82] Graves, J A M and Gecz, J and Hameister, H, *Evolution of the human X–a smart and sexy chromosome that controls speciation and development*, Cytogenet Genome Res, vol. 99(1-4), p. 141-45, 2002.

[83] Delbridge, Margaret L and McMillan, Daniel A and Doherty, Ruth J and Deakin, Janine E and Graves, Jennifer A Marshall, *Origin and evolution of*

candidate mental retardation genes on the human X chromosome (MRX), BMC Genomics, vol. 9, p. 65, Feb 2008.

[84] Splinter, Erik and de Wit, Elzo and Nora, Elphege P and Klous, Petra and van de Werken, Harmen J G and Zhu, Yun and Kaaij, Lucas J T and van Ijcken, Wilfred and Gribnau, Joost and Heard, Edith and de Laat, Wouter, *The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA*, Genes Dev, vol. 25(13), p. 1371-83, Jul 2011.

[85] Voon, Hsiao P J and Hughes, Jim R and Rode, Christina and De La Rosa-Velzquez, Inti A and Jenuwein, Thomas and Feil, Robert and Higgs, Douglas R and Gibbons, Richard J, *ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes*, Cell Rep, vol. 11(3), p. 405-18, Apr 2015.

[86] Thorvaldsen, J L and Duran, K L and Bartolomei, M S, *Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2*, Genes Dev, vol. 12(23), p. 3693-702, Dec 1998.

[87] Bell, A C and Felsenfeld, G, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene*, Nature, vol. 405(6785), p. 482-5, May 2000.

[88] Law, Martin J and Lower, Karen M and Voon, Hsiao P J and Hughes, Jim R and Garrick, David and Viprakasit, Vip and Mitson, Matthew and De Gobbi, Marco and Marra, Marco and Morris, Andrew and Abbott, Aaron and Wilder, Steven P and Taylor, Stephen and Santos, Guilherme M and Cross, Joe and Ayyub, Helena and Jones, Steven and Ragoussis, Jiannis and Rhodes, Daniela and Dunham, Ian and Higgs, Douglas R and Gibbons, Richard J, *ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner*, Cell, vol. 143(3), p. 367-78, Oct 2010.

[89] Georgel, Philippe T and Horowitz-Scherer, Rachel A and Adkins, Nick and Woodcock, Christopher L and Wade, Paul A and Hansen, Jeffrey C, *Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation*, J Biol Chem, vol. 278(34), p. 32181-8, Aug 2003.

[90] , Nikitina, Tatiana and Shi, Xi and Ghosh, Rajarshi P and Horowitz-Scherer, Rachel A and Hansen, Jeffrey C and Woodcock, Christopher L, *Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin*, Mol Cell Biol, vol. 27(3), p. 864-77, Feb 2007.

[91] Skene, Peter J and Illingworth, Robert S and Webb, Shaun and Kerr, Alastair R W and James, Keith D and Turner, Daniel J and Andrews, Rob and Bird, Adrian P, *Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state*, Mol Cell, vol. 37(4), p. 457-68, Feb 2010.

[92] Kernohan, Kristin D and Vernimmen, Douglas and Gloor, Gregory B and Berube, Nathalie G, *Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping*, Nucleic Acids Res, vol. 42(13), p. 8356-68, Jul 2014.

[93] Baumann, Claudia and De La Fuente, Rabindranath, *ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells*, Chromosoma, vol. 118(2), p. 209-22, Apr 2009.

[94] Sarma, Kavitha and Cifuentes-Rojas, Catherine and Ergun, Ayla and Del Rosario, Amanda and Jeon, Yesu and White, Forest and Sadreyev, Ruslan and Lee, Jeannie T, *ATRX directs binding of PRC2 to Xist RNA and Polycomb targets*, Cell, vol. 159(4), p. 869-83, Nov 2014.

[95] Bagheri-Fam, Stefan and Argentaro, Anthony and Svingen, Terje and Combes, Alexander N and Sinclair, Andrew H and Koopman, Peter and Harley, Vincent R, *Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome*, Hum Mol Genet, vol. 20(11), p. 2213-24, Jun 2011.

[96] Shioda, Norifumi and Yabuki, Yasushi and Yamaguchi, Kouya and Onozato, Misaki and Li, Yue and Kurosawa, Kenji and Tanabe, Hideyuki and Okamoto, Nobuhiko and Era, Takumi and Sugiyama, Hiroshi and Wada, Takahito and Fukunaga, Kohji, *Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome*, Nat Med, vol. 24(6), p. 802-813, Jun 2018.

[97] Cocquet, Julie and Ellis, Peter J I and Yamauchi, Yasuhiro and Riel, Jonathan M and Karacs, Thomas P S and Rattigan, Aine and Ojarikre, Obah A and Affara, Nabeel A and Ward, Monika A and Burgoyne, Paul S, *Deficiency in the multicopy Sycp3-like X-linked genes Slx and Slxl1 causes major defects in spermatid differentiation*, Mol Biol Cell, vol. 21(20), Oct 2010.

[98] Cocquet, Julie and Ellis, Peter J I and Yamauchi, Yasuhiro and Mahadevaiah, Shantha K and Affara, Nabeel A and Ward, Monika A and Burgoyne, Paul S, *The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis*, PLoS Biol, vol. 7(1), Nov 2009.

[99] Jacquemont, Sebastien and Coe, Bradley P and Hersch, Micha and Duyzend, Michael H and Krumm, Niklas and Bergmann, Sven and Beckmann, Jacques S and Rosenfeld, Jill A and Eichler, Evan E, *A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders*, Am J Hum Genet, vol. 94(3), p. 415-25, Mar 2014.

[100] Quinodoz, Sofia A and Ollikainen, Noah and Tabak, Barbara and Palla, Ali and Schmidt, Jan Marten and Detmar, Elizabeth and Lai, Mason M and Shishkin, Alexander A and Bhat, Prashant and Takei, Yodai and Trinh, Vickie

and Aznauryan, Erik and Russell, Pamela and Cheng, Christine and Jovanovic, Marko and Chow, Amy and Cai, Long and McDonel, Patrick and Garber, Manuel and Guttman, Mitchell, *Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus*, Cell, vol. 174(3), p. 744-57, Jul 2018.

[101] Cubelos, Beatriz and Sebastian-Serrano, Alvaro and Beccari, Leonardo and Calcagnotto, Maria Elisa and Cisneros, Elsa and Kim, Seonhee and Dopazo, Ana and Alvarez-Dolado, Manuel and Redondo, Juan Miguel and Bovolenta, Paola and Walsh, Christopher A and Nieto, Marta, *Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex*, Neuron, vol. 66(4), p. 523-35, May 2010.

[102] Hagege, Helene and Klous, Petra and Braem, Caroline and Splinter, Erik and Dekker, Job and Cathala, Guy and de Laat, Wouter and Forne, Thierry, *Quantitative analysis of chromosome conformation capture assays (3C-qPCR)*, Nat Protoc, vol. 2(7), p. 1722-33, 2007.

[103] Stadhouders, Ralph and Kolovos, Petros and Brouwer, Rutger and Zuin, Jessica and van den Heuvel, Anita and Kockx, Christel and Palstra, Robert-Jan and Wendt, Kerstin S and Grosveld, Frank and van Ijcken, Wilfred and Soler, Eric, *Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions*, Nat Protoc, vol. 8(3), p. 509-24, Mar 2013.

[104] Ferraiuolo, Maria A and Sanyal, Amartya and Naumova, Natalia and Dekker, Job and Dostie, Josee *From cells to chromatin: capturing snapshots of genome organization with 5C technology*, Methods, vol. 58(3), p. 255-67, Nov 2012.

[105] Lajoie, Bryan R and van Berkum, Nynke L and Sanyal, Amartya and Dekker, Job, *My5C: web tools for chromosome conformation capture studies*, Nat Methods, vol. 6(10), p. 690-1, Oct 2009.

[106] Splinter, Erik and de Wit, Elzo and van de Werken, Harmen J G and Klous, Petra and de Laat, Wouter, *Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation*, Methods, vol. 58(3), p. 221-30, Nov 2012.

[107] Raviram, Ramya and Rocha, Pedro P and Muller, Christian L and Miraldi, Emily R and Badri, Sana and Fu, Yi and Swanzey, Emily and Proudhon, Charlotte and Snetkova, Valentina and Bonneau, Richard and Skok, Jane A, *4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments*, PLoS Comput Biol, vol. 12(3), e1004780, Mar 2016.

[108] Yu, Guangchuang and Wang, Li-Gen and He, Qing-Yu, *ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization*, Bioinformatics, vol. 31(14), p. 2382-3, Jul 2015.